Directed molecular evolution by gene conversion

Davies, Bryn Gethin

Award date:
2001

Awarding institution:
University of Bath

DIRECTED MOLECULAR

EVOLUTION BY

GENE CONVERSION

submitted by Bryn Gethin Davies

for the degree of PhD

of the University of Bath

2001

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

No defects, only design

SUMMARY

Two projects have been carried out, both using *Saccharomyces cerevisiae* (baker's yeast) as the experimental host organism.

The first, and main project (Chapters 1 and 3), was the construction of a novel *in vivo* system for performing directed molecular evolution. At present, evolution of proteins/enzymes towards a desired functional goal is carried out *in vitro*, the favoured technique for doing this being DNA shuffling (described in Section 1.2). However, although powerful, DNA shuffling and other related techniques require numerous DNA manipulation steps and the use of expensive reagents. The aim of this project was to overcome these limitations by the creation of a cell-based system designed to mimic natural processes for the creation of diversity (e.g. antibody production in chickens), that, once constructed, would allow repeated cycles of selection against a specified target. The strategy uses the inherent gene conversion mechanism of *S. cerevisiae* - which is usually used to change mating-type - in order to effect 'DNA shuffling' between two (in the first instance) heterologous genes, integrated at synthetic loci in the yeast genome. The genes chosen for shuffling, *OR17-40* and *OR17-228*, putative human odorant receptors, represent ideal substrates; they are homologous (>80% identical at the nucleotide level); are involved in molecular recognition (could be evolved to recognize a molecule of interest); and are G-proteins (allowing them to be coupled, via a signal transduction pathway, to a reporter gene). Two strains containing the experimental model system (Section 1.3) were constructed, one strain being a mismatch repair defective mutant of the other (for reasons described in Section 1.3). These strains were designed to look at the efficacy of the *in vivo* 'DNA shuffling' reaction (the odorant receptor genes were not coupled to a reporter gene at this stage). Also constructed were two control strains for looking at different aspects of the experimental reaction. Preliminary experiments are reported for all the strains made. In the experimental strains the desired gene conversion reaction was not observed. Reasons for this are discussed in light of evidence from the control strains that I-*Sce*I - the endonuclease responsible for initiating the reaction - was not being expressed. Possible remedies to this problem are suggested.

The second, subsidiary, project was to investigate the efficacy of a Cre-based double recombination system to target DNA site-specifically into the yeast

genome (Chapter 4). The model system used for this project comprises of an antibiotic resistance marker (the *kanMX4* module) integrated at the chromosomal location to be targeted (*leu2*) and a plasmid carrying the gene to be positioned at the locus (*OR17-228*). Both the *OR17-228* gene and the *kanMX4* module are flanked on one side by a *loxP* site and on the other by a *loxP 511* site. With this model system it is shown that a single copy of the *OR17-228* gene can be targeted to the *leu2* locus in a little over 20% of yeast cells induced to undergo the exchange. Analysis of several targeted loci shows that the integrated gene is in the intended orientation, and that there are no ectopic copies present. Ways in which the system might be optimised are suggested, with the hope that it will provide a valuable new method for creating integrated repertoires in yeast.

ACKNOWLEDGEMENTS

Firstly, and foremost, I would like to thank my supervisor, Dr. Jonathan Cox, for the opportunity to do this project, and for his guidance and support throughout.

I am very grateful to Dr. Alan Wheals for his help and advice on various yeast related technical issues. To Dr. Andrew Ward for use of laboratory space to carry out radioisotope work and to Bill Bennett and Derek Paisley for assistance with this work.

I would also like to thank the BBSRC for funding, and F. Winston/L. Pacella, S. G. Oliver/N. S. Zhang, A. E. Wheals/T. Edwards, and B. Dujon who generously supplied strains FY10/FY22/FY833, the pRS vectors, vector pFA6a-kanMX4 and vector pPEX7, respectively.

Finally I would like to say a special thank you to Saima Aijaz for her contribution to this work and to all members of laboratory 0.34 past and present for creating a happy and friendly workplace.

ABBREVIATIONS

General

A_{600}	absorbance at 600 nm
ARS	autonomously replicating sequence
b.p.	base pairs
CEN	centromere
CHO	Chinese hamster ovary (cell line)
Ci	Curie
dsDNA	double stranded DNA
DSBR	double strand break repair
k.b.p.	kilobase pairs
LMP	low melting point
MCS	multiple cloning site
n.t.	nucleotide
ORF	open reading frame
ORI	origin of replication
PCR	polymerase chain reaction
r.p.m.	revolutions per minute
SC	synthetic complete
SGD	*Saccharomyces cerevisiae* genome database
SOE	splicing by overlap extension
UV	ulta-violet
v/v	ratio of volume to volume
w/v	ratio of weight to volume
YPD	yeast peptone dextrose
Δ	deletion
::	novel junction (fusion or insertion)

Chemical

BSA	bovine serum albumin
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate

dGTP	deoxyguanosine 5'-triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside 5' triphosphate
dTTP	deoxythymidine 5'-triphosphate
EDTA	disodium ethylenediamine tetraacetate
G418	geneticin
GLU	glucose
GAL	galactose
mRNA	messenger RNA
ONPG	o-nitrophenyl-β-D-galactopyranoside
PEG	polyethylene glycol
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
TAE	40 mM Tris-acetate, 1 mM EDTA
TBE	90 mM Tris-borate, 2 mM EDTA
Taq	thermostable DNA polymerase from *Thermus aquaticus*
Tris	tris-(hydroxymethyl)-methylamine

Amino acids

A	Ala	Alanine	L	Leu	Leucine
R	Arg	Arginine	K	Lys	Lysine
N	Asn	Asparagine	M	Met	Methionine
B	Asp	Aspartate	F	Phe	Phenylalanine
C	Cys	Cysteine	P	Pro	Proline
E	Glu	Glutamate	S	Ser	Serine
Q	Gln	Glutamine	T	Thr	Threonine
G	Gly	Glycine	W	Trp	Tryptophan
H	His	Histidine	Y	Tyr	Tyrosine
I	Ile	Isoleucine	V	Val	Valine

Bases

A	Adenine	T	Thymine
C	Cytosine	G	Guanine

TABLE OF CONTENTS

CHAPTER 5 TOWARDS A CRE-BASED RECOMBINATION SYSTEM FOR GENERATING INTEGRATED DNA REPERTOIRES SITE-SPECIFICALLY IN YEAST

CHAPTER 1

INTRODUCTION

There are a number of important areas where the ability to rapidly evolve a protein in the laboratory would be extremely beneficial, e.g. therapy, diagnostics and biosensors. Probably the best way to evolve a protein is by homologous recombination in conjunction with a small amount of random mutation. DNA shuffling ('sexual PCR') is a very clever *in vitro* method which achieves just this (reviewed in Section 1.2). The technique entails digesting the gene encoding the protein of interest with DNaseI and then reassembling the gene in a primerless 'inverse chain reaction', in which the homologous pieces of DNA prime on themselves. During reassembly a small number of mutations may be introduced as a result of the infidelity of the polymerase. The recombinant genes are transformed into bacteria and selection performed. The genes of improved mutants are collectively treated to 2 - 4 further rounds of digestion/assembly/transformation and selection. In these additional rounds, advantageous mutations from different mutants may be combined to give new, vastly improved mutants. It is because of this combinatorial potential that DNA shuffling is so powerful. However, despite the evident efficacy of the process, it suffers from a number of disadvantages, notably the expense of the reagents and having to extract and retransform the DNA during each round. In certain cases, e.g. where the usefulness of a mutant repertoire created by shuffling was such that it was required for a variety of different selection purposes, these disadvantages may prove serious. In this study an alternative recombination/protein evolution strategy is attempted which avoids these disadvantages. The strategy relies on **gene conversion**.

Gene conversion

Gene conversion can be simply defined as the conversion of one gene's sequence to that of another (Weaver and Hedrick, 1992). This occurs in a non-reciprocal manner, such that the donor gene remains unaffected. The effects of gene conversion are observed during analysis of fungal recombination. When fungi such as *Neurospora crassa* (pink bread mould) sporulate, two haploid nuclei fuse, producing a

diploid nucleus that undergoes meiosis to give four haploid nuclei. These nuclei then experience mitosis to produce eight haploid nuclei, each appearing in a separate spore. In principle, if one of the original nuclei contained one allele (*A*) at a given locus, and the other contained another allele (*a*) at the same locus, differing by a single point mutation, then the mixture of alleles in the spores should be equal: four *A*'s and four *a*'s (corresponding to a segregation ratio of 4:4). However, due to postmeiotic segregation and gene conversion events (shown diagrammatically in Figure 1.1) aberrant ratios (3:5 and 2:6) are observed 0.1% to 1% of the time (depending on fungal species).

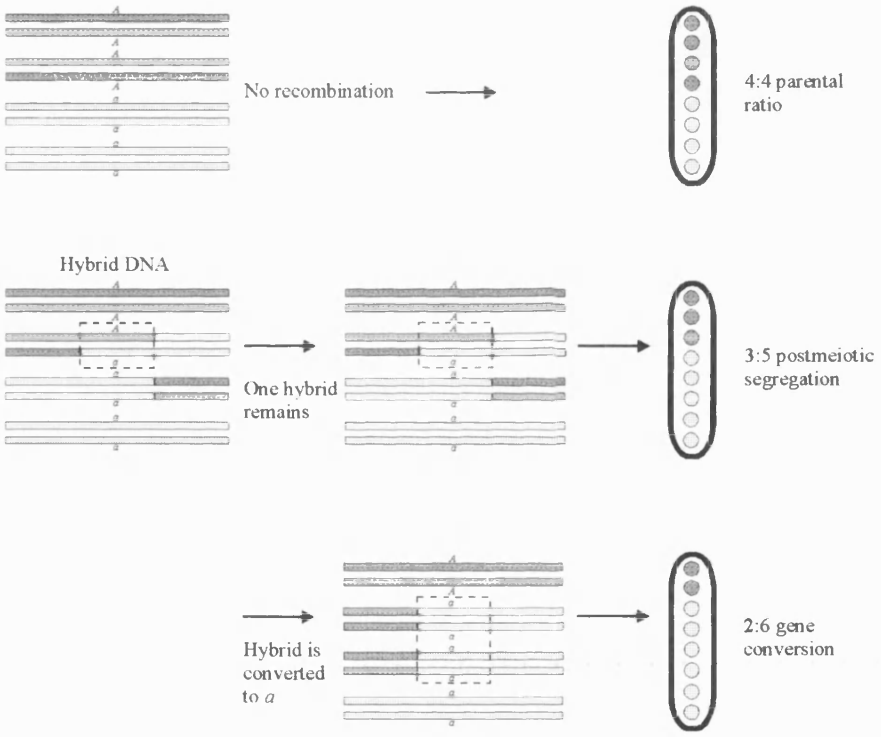


Figure 1.1 Spore formation in the fungus *Neurospora*. The four copies of each allele following meiosis are shown (top left). Note that each chromatid (represented by 'railroad track') contains two copies of each allele (one on either DNA strand). Following recombination, mitosis results in each individual strand (8 in total) being duplicated, giving the spores shown (RHS). Diagram reproduced from Lewin (1994).

Gene conversion events can apply to the conversion of a single point mutation, as in Figure 1.1, or to the conversion of a whole gene or DNA tract (of most interest here), as described below.

The protozoan flagellate parasite *Trypanosoma brucei* is the paradigmatic species of African trypanosomes (Payes and Nolan, 1998). Transmitted by tsetse fly, it causes sleeping sickness in humans (more than 50,000 death cases per year) and nagana disease in cattle (as a consequence of the disease 10 million square kilometres of potential grazing land, in 37 countries of sub-Saharan Africa, are rendered unsuitable for livestock breeding). The devastating effect of the pathogen lies in its ability to evade its host's immune system by repeatedly changing its surface coat, the only parasite structure exposed to host antibodies (Borst and Greaves, 1987). The surface coat consists mainly of a single protein species, the variant surface glycoprotein (VSG). The frequency with which the new coat arises ranges from 10^{-2} to 10^{-6} per division (Van der Ploeg *et al.*, 1992). In a similar manner, another disease causing organism, the bacterium *Borrelia hermsii* (which causes relapsing fever in humans), also relies on antigenic variation to escape from host antibodies (Restrepo and Barbour, 1994). In *Borrelia* the immunodominant surface protein is called the variable major protein (VMP), the appearance of new serotypes occurs at an estimated frequency of 10^{-4} per division (Stoenner *et al.*, 1982). In both organisms the change in the surface protein is associated with rearrangements in the genome driven by gene conversion mechanisms, whereby a functional gene - expressing the surface protein - is converted into a different functional gene. This occurs by transfer of information from one or more homologous silent donor genes or pseudogenes (Figure 1.2).

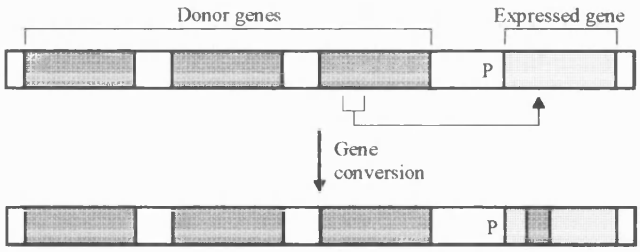


Figure 1.2 Generalised representation of the process of antigen switching in *T. brucei* and *B. hermsii*. Potential donors for templated gene conversion of the functional gene (light grey rectangle) are located upstream of the expression site as a pseudogene cluster (dark rectangles). Only the gene at the expression site is expressed due to the presence of the promoter (P) element. Only 3 potential donor genes are shown, although (for trypanosomes) this number may represent 1000 or more. Gene conversion, resulting in a different surface protein being expressed, may result from two processes: **1**. the entire expression linked gene being replaced with a silent copy, or **2**. in parts of the gene being replaced (as shown) by information from one or more donor genes, resulting in a 'mosaic' gene.

By creating 'mosaic' genes, both *T. brucei* and *B. hermsii* increase the diversity further of the surface proteins they are capable of expressing. It is interesting that this is also the method employed in the avian immune system in order to generate antibody diversity. Reynaud and co-workers (1987) demonstrated that chickens follow a route to immunoglobulin production that differs radically from that found in mammals (mammalian antibody production is reviewed by Alt *et al.*, 1992). In chickens there is only a single functional V-J-C region for the lambda light chain of the immunoglobulin, which are joined in a precise way. Upstream of the functional $V_{\lambda 1}$ gene lie 25 V_{λ} pseudogenes (lacking either the coding segment at one or both ends, or a proper signal for recombination) organised in either orientation. Diversity in the light chain genes is created solely by somatic mutations, brought about by multiple rounds of segmental gene conversion (analogous to that shown in Figure 1.1), in which the segments (between 10 and 120 b.p. in length) are derived from the pseudo-V genes. Gene conversion takes place at a frequency of 0.05 - 0.1 per cell generation. At the end of the immune maturation period, a rearranged $V_{\lambda 1}$ sequence has 4 - 6 converted segments spanning its entire length, derived from different donor pseudogenes; this allows (if all pseudogenes participate) 2.5×10^8 combinations.

Experimental Aim

If the same process used to generate antigen and antibody variation (above) could be mimicked in a suitable host organism, using the inherent recombination machinery of that organism to effect gene conversion at a synthetic donor/acceptor gene locus, the disadvantages of DNA shuffling would be overcome.

The possibility of imitating the process of gene conversion (adopted by birds) in order to generate large repertoires was first proposed by Winter and Milstein (1991).

It is the aim of this project to create such an '*in vivo* DNA shuffling system' using the host organism *Saccharomyces cerevisiae* (baker's yeast), since is genetically tractable and since in the mating-type switch it already has an operational (and well characterised) gene conversion mechanism. Furthermore, homologous recombination in this organism is extremely efficient (Orr-Weaver *et al.*, 1981).

The proposed donor/acceptor locus and the experimental strategy are described in detail in Section 1.3. The host organism, gene conversion therein, and a brief review of *in vitro* DNA shuffling are presented next.

1.1 THE HOST ORGANISM: *Saccharomyces cerevisiae*

Saccharomyces cerevisiae, a unicellular fungus, is one member of a genus containing some forty species, each of which produces spherical to ellipsoidal cells by budding, produce ascospores in asci and are capable of the efficient conversion of sugars to alcohol (Berry, 1982). Initially the name *Saccharomyces* was applied to all yeasts isolated from alcoholic beverages, three species were recognised by Meyen (1837) by their origin; *Saccharomyces vini* from wine, *S. cerevisiae* from beer and *S. pomorum* from cider (in 1870 the genus *Saccharomyces* was restricted to those yeast which produce spores). The most important *Saccharomyces* species in brewing was isolated in 1881 by the micro-biologist Emil Hansen, and described as *Saccharomyces cerevisiae* var *ellipsoideus*. Strains classified as *S. cerevisiae* have been widely used for brewing, distilling, wine making, the production of baker's yeast and biomass production.

By virtue of several of its features, *Saccharomyces cerevisiae* has become an important experimental organism for the study of eukaryotic molecular genetics. Because it divides by budding, it is easy to follow the progress of the cell cycle in a living cell (Murray and Hunt, 1993). It is able to propagate happily in either the haploid or the diploid state (described further below), so the effects of cell type or ploidy on gene activity can be determined. It can be grown on defined media, allowing the investigator complete control of its chemical and physical environment. Further, it has powerful and fast genetics, making it easy to isolate mutants in cellular processes and to clone the genes identified. Specific genes of interest can be deleted or overexpressed at will.

The complete DNA sequence of *Saccharomyces cerevisiae* chromosome III was produced in 1992, representing the first entire chromosome from any organism to be analysed (Oliver *et al.*, 1992). Four years later, as a result of an international effort, the sequence of the entire genome was published (Goffeau *et al.*, 1996; Goffeau *et al.*, 1997); making *Saccharomyces cerevisiae* the first eukaryotic organism to have its genome completely sequenced. The ~13.5 M.b. sequence revealed, distributed somewhat unevenly amongst yeast's 16 haploid chromosomes, approximately 6000 genes (average size 1450 b.p.). Features of the genome include, a high level of compactness (a protein encoding gene is found for every 2 k.b.p. of the yeast genome, in comparison to every 6 k.b.p. for the nematode worm and every 30 k.b.p. for the

human sequence), the presence of few introns and little repetitive DNA. The sequence also showed that the yeast genome contained significant redundancy, with 53 duplicated gene clusters representing more than 30% of the genome (Clayton *et al.*, 1997). Further, of the genes identified, about 2000 represent orphan genes, having no known function or structural homologues of known function; it is noted by B. Dujon (1996) that the sequence has served to highlight the great extent of what remains to be understood, in an otherwise extensively studied organism. Determining the biological role of such orphan genes is the focus of international efforts, such as the EUROFAN (European Functional Analysis Network) project, which aims to study 1000 of the unknown genes, by the creation of deletion strains.

The complete *S. cerevisiae* DNA sequence is freely available on the internet, from the *S. cerevisiae* Genome Database (SGD), heralding an exciting new 'post-genomic' era in yeast genetics and research. One that, using powerful new technologies, such as DNA chips (oligonucleotide arrays) [Pease *et al.*, 1994; Che *et al.*, 1996; Ramsey, 1998], will allow us to simultaneously monitor the expression level of every gene in the yeast genome (Wodicka *et al.*, 1997). Providing previously unimaginable amounts of information about gene function and the interrelationship of genes in biological pathways.

1.1.1 Gene conversion in *Saccharomyces cerevisiae*: the mating-type switching reaction

Barring natural cellular events, gene conversion is not used by *Saccharomyces cerevisiae* to generate diversity within its genome, as for, e.g. *B. hermsii*. Instead it is used to effect a change in the sex, or mating-type, of a cell. The mating-type locus of *S. cerevisiae* (one of the first genetic loci in yeast to be studied), the genetic rearrangements at the locus, and the genes expressed thereof, have been the subjects of extensive research; for reviews see e.g., Haber (1983), Haber (1998a). The brief review presented below is intended only to describe the most salient points of the system, and that work of relevance to this project.

The alternation of haploid and diploid phases in the life cycle of yeast was established by Winge and Hjort (1935). Lindegren and Lindegren (1943) isolated stable haploids that could be distinguished by their ability to mate with each other and form non-mating diploids. The two mating types, designated **a** and **α**, were

determined by two alleles of a single genetic locus, *MAT***a** and *MAT*α (located on the right arm of chromosome III; Mortimer and Hawthorne, 1969). The two alleles were found to be co-dominant, because the *MAT***a**/*MAT*α diploids formed by conjugation had a distinctly different phenotype from either haploid; diploids were nonmating. When placed under nitrogen starvation conditions in the presence of acetate (non-fermentable carbon source), *MAT***a**/*MAT*α diploids undergo meiosis and form an ascus containing four haploid spores, two *MAT*α and two *MAT***a** (Roman and Sands, 1953). The complete life cycle of yeast (reflecting also later discoveries), is shown in Figure 1.3.

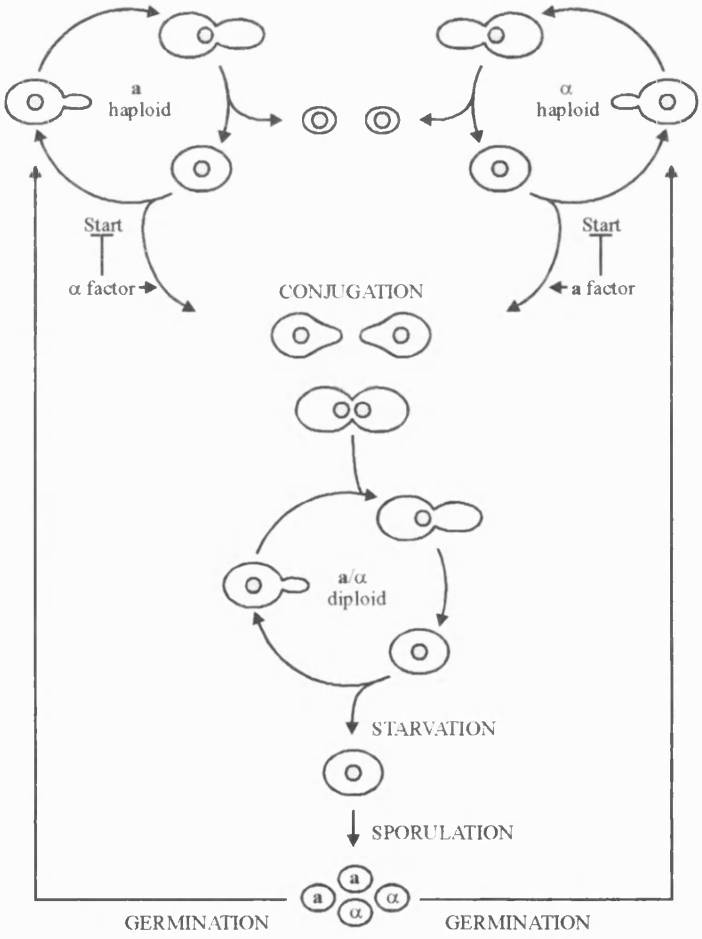


Figure 1.3 The life cycle of *S. cerevisiae*. Yeast can reproduce asexually by budding, as either haploid or diploid cells. Haploid cells exist in two mating types, **a** and α, and cells of opposite mating-types can mate to form diploid cells. Haploid budding cells secrete mating factors that bind to receptors on each other's surface. This chemical signalling induces cell cycle arrest and the expression of genes that carry out the physical process of mating. Under starvation conditions, the diploid cells sporulate to yield four haploid spores. Spore germination results in haploid budding cells that can mate with each other to restore the diploid state. Diagram reproduced form Murray and Hunt (1993).

Producing spores of different mating-type within a single ascus ensures that when the spores germinate, those of opposite mating-type have a good chance of mating, restoring the diploid state. The diploid state provides protection from the burden of harmful mutations (Resnick and Martin, 1976; Chlebowicz and Jachymczyk, 1979; Fairhead and Dujon, 1993), and allows rapid sporulation in response to unfavourable environmental condition. A second strategy also ensures rapid diploidization, haploid cells of one mating-type are able to switch to the other mating-type. Mating-type switching (homothallism) ensures that even if a spore germinates in isolation, it will be able to find a partner with which to mate.

Mating-type switching in homothallic yeast was studied by Strathern, Hicks and Herskowitz who deduced rules defining the pattern of switching (Hicks and Herskowitz, 1976; Strathern and Herskowitz, 1979). First, only those cells that have previously budded are capable of switching. Second, cells switch in pairs; i.e., if a cell undergoes mating-type interconversion, then both mother and daughter produced from that cell are switched. Third, a cell capable of switching will switch to the opposite mating-type around 85% of the time. Shown diagrammatically in Figure 1.4.

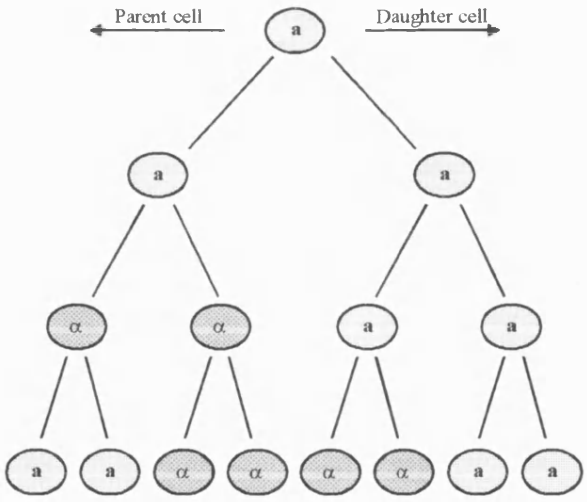


Figure 1.4 The pattern of mating-type switching as determined by Strathern, Hicks and Herskowitz.

In most laboratory strains (as developed by Lindegren and Lindegren, 1943), the mating-type alleles are essentially stable (heterothallic strains), although rare conversions of *MAT***a** to *MAT*α (or vice versa), of the order of 10^{-6} can occur (Klar *et al.*, 1982).

The frequency of switching in homothallic strains, almost every other cell division, is much more rapid than could be anticipated by any mechanism involving conventional spontaneous mutations. Initially it was proposed that the rapid switching of mating-type might be due to a flip-flop type mechanism (similar to that used by the bacteria *Salmonella* to express two different proteins within its flagella), where transcription at *MAT* could be directed to copy either **a** or **α** sequences that were both present at *MAT* (Hawthorne, 1963; Brown, 1976). Later genetic analysis, which showed that mutations at the *MAT* locus could be healed (e.g., Strathern and Herskowitz, 1979), ruled out such explanations.

An explanation of the high frequency of sex reversal came from genetic crosses, that showed that chromosome III possesses both an active *MAT* gene and two unexpressed mating-type loci, *HML* and *HMR* (Naumov and Tolstorukov, 1973; Harashima *et al.*, 1974). An **a** to **α** switch requiring *HML***α**, located on the left arm of chromosome III, and an **α** to **a** interconversion requiring *HMR***a**, located at the far end of the right arm of chromosome III. Genetic and physical analysis of the *MAT* loci revealed that interconversion occurs by a replacement of information (a cassette) present at *MAT* with a copy of information (a cassette) of the opposite mating-type present at one of the *HM* loci (Hick *et al.*, 1977; Strathern *et al.*, 1980; Nasmyth and Tatchell, 1980). In other words, *HML***α** is a silent or unexpressed copy of *MAT***α**, and *HMR***a** is a silent or unexpressed copy of *MAT***a**. This is the cassette model of mating-type switching, and is shown diagrammatically in Figure 1.5.

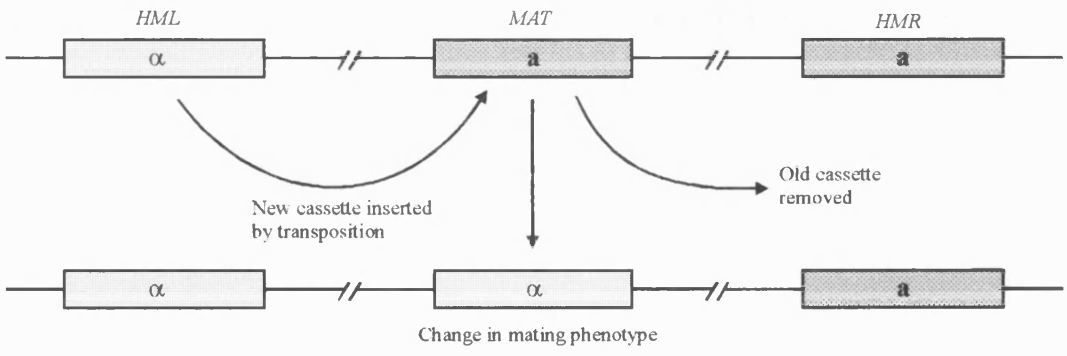


Figure 1.5 The cassette model of mating-type switching. *HML* and *HMR* contain silent mating-type information. Here the **α** cassette from *HML* replaces the **a** cassette at the expressed *MAT* locus. Most laboratory strains have the **α** information at *HML* and **a** information at *HMR* (type II arrangement) as shown; variants containing the opposite *HML***a** *HMR***α** configuration (type I) have been isolated (Klar *et al.*, 1982).

Analysis of cassette switching showed that it occurred in a non-reciprocal manner, such that the donor information at either *HML* or *HMR* replaces the information present at the *MAT* locus, but does not exchange with it. Thus, the genetic information needed to promote subsequent mating-type switches is preserved.

Cassette exchange is a directed process, dependent on the phenotype of the cell (Klar *et al.*, 1982). Switching usually involves replacement of *MAT***a** by the copy of the cassette from *HML*α, or replacement of *MAT*α by the copy of the cassette at *HMR***a**. The dependence on the cell's phenotype (**a** or α) rather than the nature of the sequence present at either of the donor loci, was shown in strains possessing the reverse *HML***a** *HMR*α arrangement, which switched phenotype inefficiently (at a maximum of ~6%), as a result of futile *MAT* locus switches, that is *MAT***a** to *MAT***a** and *MAT*α to *MAT*α. The high frequency of mating-type interconversion reported by Strathern, Hicks and Herskowitz (discussed earlier) is explained by donor preference during the switching reaction (purely random donor selection would have resulted in a maximum frequency of observable switches of 50%). The cellular mechanisms resulting in directed switching, as determined by the phenotype of the cell, are described later.

Mating-type information

The four loci, *HMR***a**, *MAT***a**, *MAT*α and *HML*α were isolated using recombinant techniques (Nasmyth and Tatchell, 1980; Hicks *et al.*, 1979). DNA heteroduplex analysis was used to define the dimensions of the loci and their physical relationship (Nasmyth and Tatchell, 1980; Strathern *et al.*, 1980). The exact dimensions (nucleotide length) of the loci were determined by Astell *et al.* (1981). Thus: the *MAT***a** and *MAT*α loci are about 2500 nucleotide pairs in length. Each contains a central unique sequence, 642 nucleotide pairs in **a** and 747 in α. These unique sequences are bound by homologous sequences of about 1400 nucleotide pairs on the left (centromere proximal) and 320 on the right. The *HML*α locus, on the left arm of chromosome III, is homologous with *MAT*α in structure; *HML***a** and *MAT***a** are also homologous. However, *HMR***a** (and *HMR*α) on the right arm of chromosome III between *MAT* and the telomere, lacks about 700 nucleotide pairs from the left end and about 90 from the right end when compared with *MAT***a** (and *MAT*α). The segments of the cassette have been designated, from left to right, W, X, Ya (Yα), Z1 and Z2,

and all are present in *HML***a** and *MAT***a** (and *HML*α and *MAT*α). Only X, Ya (Yα) and Z1 are present in *HMR***a** (and *HMR*α). Shown diagrammatically in Figure 1.6.

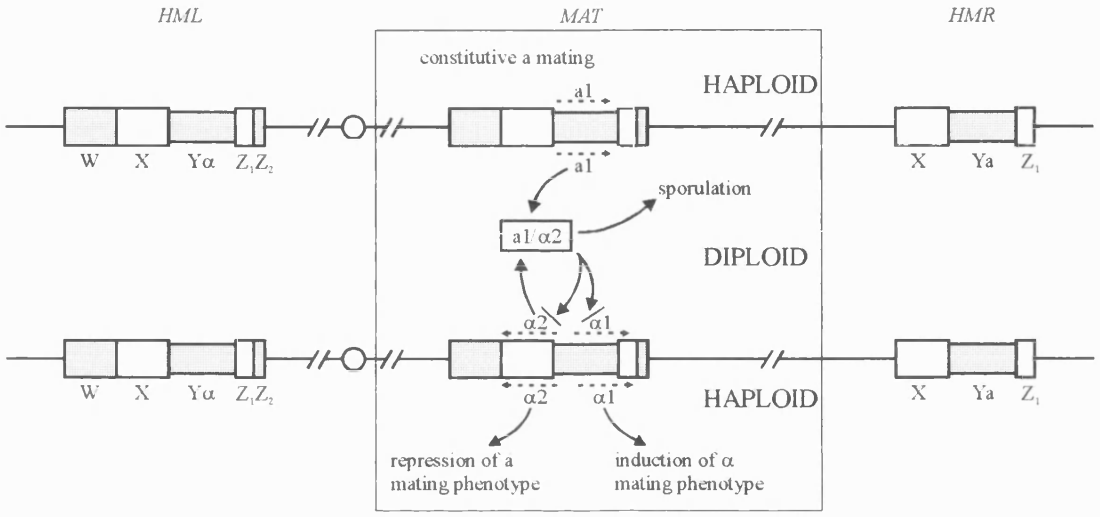


Figure 1.6 An overview of yeast mating-type loci on chromosome III and control of cell type by *MAT***a** and *MAT*α. Boxed area refers to the activity of *MAT***a** and *MAT*α in their determination of **a**, α, or **a**/α cell types. Positive controls are represented by arrows, negative control by arrows interrupted by a bar. Diagram and legend reproduced from Nasmyth *et al.* (1981).

The position and orientation of the two major transcripts produced at *MAT***a** (**a**1 and **a**2) and at *MAT*α (α1 and α2) were determined by Nasmyth *et al.* (1981). The **a**1 mRNA is encoded entirely within the **a**-specific DNA sequence (the **a**1 protein, containing 148 amino acids, results from a readthrough of a UGA at codon 45). The **a**2 mRNA, which is transcribed divergently from **a**1 mRNA, is encoded in a region common to both *MAT***a** and *MAT*α. The α1 and α2 mRNAs are also transcribed divergently and have their 5' starts about 240 nucleotides apart (centred toward the left end) within the α-specific sequence. Most of the α1 transcript is encoded in Yα, with the 3'-end encoded in the Z segment. Most of the α2 segment is encoded in the W segment.

The *MAT* locus determines cell-type by the control of unlinked **a** or α specific genes, rather than by direct expression of all structural information required for mating (Strathern *et al.*, 1981). The *MAT*α1 gene product (in conjunction with a constitutive protein, Mcm1p; Hagen *et al.*, 1993) is necessary for the expression of genes unlinked to *MAT* that specify an α mating-type (e.g., the α-factor structural gene *MF*α, and the **a**-factor receptor gene, *STE3*). *MAT*α2 is necessary (in conjunction with

Mcm1p; Keleher *et al.*, 1989) for repressing **a** specific genes, which are otherwise expressed constitutivley and would therefore antagonize the expression of an **α** mating-type. *MAT***a** plays no part in the specification of an **a** mating-type, as the genes necessary are constitutivley expressed in the absence of the **α**2 gene (e.g., the **a**-factor structural gene *MF***a**, and the **α**-factor receptor gene, *STE2*). However, the *MAT***a**1 gene is necessary for specifying, in conjunction with the **α**2 gene, the non-mating **a**/**α** diploid cell-type and the ability to sporulate (Figure 1.6). No biological role has been assigned to the *MAT***a**2 gene (Haber, 1998b).

Silencing of HML and HMR

The donor loci, *HML* and *HMR*, contain complete copies of the mating-type genes, including their promoters, but are not expressed. Regulation of the silent mating-type genes depends on *cis*-acting sequences, termed silencers (below), as well as several *trans*-acting factors, including histones H4 and H3 (Kayne *et al.*, 1988; Thomson *et al.*, 1994), the proteins Rap1p (repressor activator protein 1; Kurtz and Shore, 1991; Kyrion *et al.*, 1993) and Abf1p (autonomous replicating sequence [ARS] binding factor 1; Buchman *et al.*, 1988) and four *SIR* (silent information regulator) proteins (*SIR1-4*; Rine *et al.*, 1979; Klar *et al.*, 1979). Failure of any one of the *trans*-acting factors to act result in the constitutive expression of the **α** and **a** genes at *HML* and *HMR*.

Deletion analysis of regions adjacent to *HML* and *HMR* defined the *cis*-acting silencing elements, known as *E* and *I* (for essential and important), found approximately 1 kilobase to either side of the internal promoter regions (Abraham *et al.*, 1984; Feldman *et al.*, 1984). These sites, each less than 150 b.p., are required for regulation of *HML* and *HMR* by the *trans*-acting factors (excluding histones H4 and H3). The roles that the two silencers play in silencing the adjacent mating-type genes differs at the two loci. At *HML*, the presence of either the *E* or *I* site alone is sufficient to silence *HML***α**, whilst at *HMR*, a deletion of the *I* site does not abolish silencing but a deletion of the *E* site does.

The silencers function in an orientation-independent manner over a reasonable distance (>4000 b.p.) to promote repression on the silent mating-type loci (Brand *et al.*, 1985; Boscheron *et al.*, 1996). This repression occurs in a region-specific, rather than a gene-specific manner (Schnell and Rine, 1986; Brand *et al.*, 1985). Each of the four silencers is composed of various combinations of clearly defined elements (Shei

and Broach, 1995). The *HMR E* silencer encompasses distinct recognition sites for the DNA binding proteins Rap1p and Abf1p as well as an ARS consensus recognized by the origin recognition complex (ORC). Deletion of any one of these elements has little effect on repression, while deletion of any two causes complete derepression. Similarly, both *HML E* and *HML I* silencers each encompass two of these three elements: *HML E* contains a Rap1p binding site and an ARS element, and *HML I* contains an Abf1p binding site and an ARS element. Thus, the silencers can be defined as collections of any two of three well-defined elements, shown in Figure 1.7.

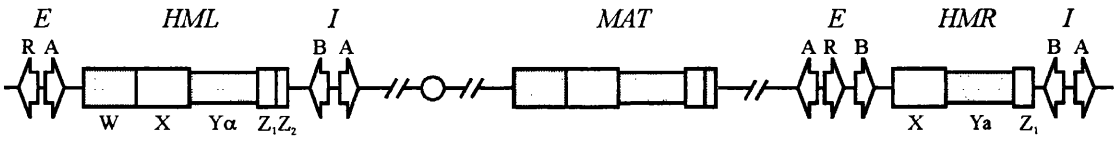


Figure 1.7 Positions of the four silencers with respect to the *HM*-loci, and organisation of the elements within the silencers. B, R, and A refer respectively, to Abf1p binding site, Rap1p binding site, and ARS core consensus sequences. Reproduced from Shei and Broach (1995).

Silencing of the *HM* mating-type loci (also subtelomeric regions) is caused by the localised formation of an altered chromatin structure (Triolo and Sternglanz, 1996). The first suggestion that chromatin structure might be important in repression came from experiments by Nasmyth (1982), comparing the DNaseI digestion profiles of *HML*, *HMR*, and *MAT* in Sir^+ and Sir^- strains. A strong *SIR*-dependent difference in the DNaseI nuclease sensitivity pattern of *HML* and *HMR* was seen, indicating that the placement of nucleosomes might play a role in repression. A causal connection between changes in chromatin structure and gene expression in *S. cerevisiae* came from the demonstration that amino-terminal deletion mutations encoded by the gene for histone H4, that removed much of the conserved basic region, led to the derepression of the *HM* loci (Kayne *et al.*, 1988). Much evidence has now been gathered, relating to accessibility of the silent loci to various DNA modifying agents and acetylation levels of histones H3 and H4 (hypo-acetylated), suggesting that transcriptional silencing derives from a heterochromatin-like organisation of these loci (e.g. Gottschling *et al.*, 1992; Braunstein *et al.*, 1993). By means of high-resolution micrococcal nuclease mapping, Weiss and Simpson (1998) and Ravindra *et al.* (1999), demonstrated directly the presence of continuous arrays of well positioned nucleosomes (closely packed dimers with a linker of <5 b.p.) spanning both *HMLα*

and *HMR***a**, in contrast to the random organisation of *MAT***a** and regions adjacent to the silencers.

Establishment of modified chromatin at the *HM* mating-type loci begins at 'initiation sites', which correspond to the *E* and *I* silencers, and spreads across the loci (Hecht *et al.*, 1995). Modified chromatin is achieved through the binding of Sir3p and Sir4p to histones H3 and H4, which results in a highly positioned array of nucleosomes. The *E* and *I* silencers serve to seed the formation of a Sir3p-Sir4p array by the binding of Sir3p to Rap1p (Moretti *et al.*, 1994) and of Sir4p to the origin recognition complex via Sir1p (Triol and Sternglanz, 1996). It is thought that Sir2p, which has been shown to interact with Sir4p, may help to stabilise the Sir3p-Sir4p array, also to play a role in regulating the level of histone hypoacetylation (Moazed *et al.*, 1997; Strahl-Bolsinger *et al.*, 1997). The onset of repression has been shown to be strongly associated with the cell cycle and replication of DNA (Miller and Nasmyth, 1984; Bi and Broach, 1997). A model showing the interactions of the various *trans*-acting factors is shown in Figure 1.8.

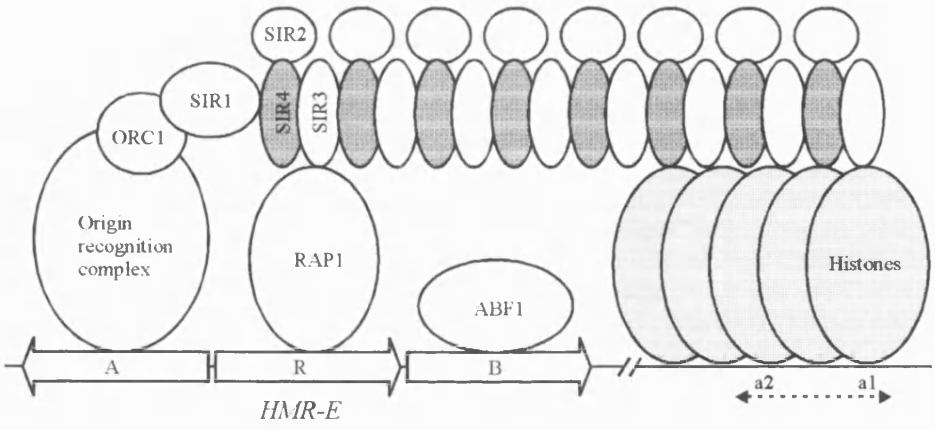


Figure 1.8 A model for the establishment of silencing at *HMR-E*. Seeding of the Sir3p-Sir4p array begins at the silencer and spreads to silence nearby genes. B, R, and A are as described in Figure 1.7. The role of Abf1p in silencing was demonstrated by Loo *et al.* (1995), although its interactions in this model are unknown and are therefore ignored. Adapted from Triolo and Sternglanz (1996).

Molecular mechanism of MAT switching

The conversion of one mating-type to another involves the replacement at the *MAT* locus of Y**a** with Y**α**, or vice versa. Evidence for physical pairing of the loci (as opposed to a replacement event mediated via a mobile genetic element) came from observations of reciprocal recombination events between *MAT* and *HMR* in

*mat***a***1*/*MAT*α*-inc* diploids (which are α-mating, due to the *mat***a***1* mutation, and undergo switching) [Haber *et al.*, 1980b]. The formation of *MAT*/*HMR* deletions during mating-type switching argued that the donor (*HMR* or *HML*) and recipient (*MAT*) sequences must physically pair during the transposition event. The switching process is initiated by a double-strand cut made in the recipient *MAT* locus by an endonuclease encoded by the *HO* gene (Strathern *et al.*, 1982; Kostriken *et al.*, 1983; Kostriken and Heffron, 1984). The HO endonuclease recognises a 24 b.p. DNA sequence just to the right of the Y-Z1 junction (Nickoloff *et al.*, 1986). Such breaks are only made in homothallic (HO^+) strains, heterothallic (HO^-) strains do not switch mating-type because they fail to make this enzyme.

The pattern of mating-type switching deduced by Strathern, Hicks and Herskowitz (Figure 1.4) is explained by the regulation of the HO endonuclease. In homothallic yeast, endonuclease activity is limited to the G1-phase of the cell cycle and restricted to haploid mother cells that have divided at least once; a single switched cell gives rise to two progeny cells with the new mating-type, indicating that the switch occurs prior to replication of the *MAT* locus (Nasmyth, 1983). The *HO* gene is under the control of *cis*-acting sites found in the 1500 b.p. upstream of the gene (Nasmyth and Shore, 1987). The general pattern of control is that repression at any one of many sites, responding to several regulatory circuits, may prevent transcription of *HO*. Mating-type control resembles that of other haploid-specific genes, transcription is prevented (in diploids) by the **a**1/α2 repressor.

The conversion process (Y**a**→Yα, Yα→Y**a**) is highly directional, in that the sequences at *MAT* are replaced by copying new sequences from either *HML*α or *HMR***a**, while the two donor loci remain unchanged by the transaction (Haber, 1998b). Directional gene conversion reflects the fact that HO endonuclease cannot cleave its recognition sequence at either *HML* or *HMR*, as these sites are apparently occluded by nucleosomes in silent DNA (Nasmyth, 1982; Weiss and Simpson, 1998).

In the experiments of Orr-Weaver *et al.* (1981) the stimulation of recombination between plasmids and a homologous chromosomal site by a double strand DNA break (or gap) was investigated. In their experiments, an internal segment of a cloned yeast gene was excised and the plasmid not religated. When this broken, gapped DNA was used in a transformation experiment, it was repaired by a gene-conversion event involving an intact homologous sequence on the chromosome. The initiation by the HO endonuclease of a double-strand break at *MAT* creates a

somewhat analogous situation in which homologous sequences are interrupted, not by a gap, but by nonhomologous sequences. Early models (e.g. Watson *et al.*, 1987) for repair events at *MAT* were based on the experiments of Orr-Weaver *et al.* (1981) and the mechanism of double-strand break repair proposed by Szostak *et al.* (1983); such a model is shown in Figure 1.9.

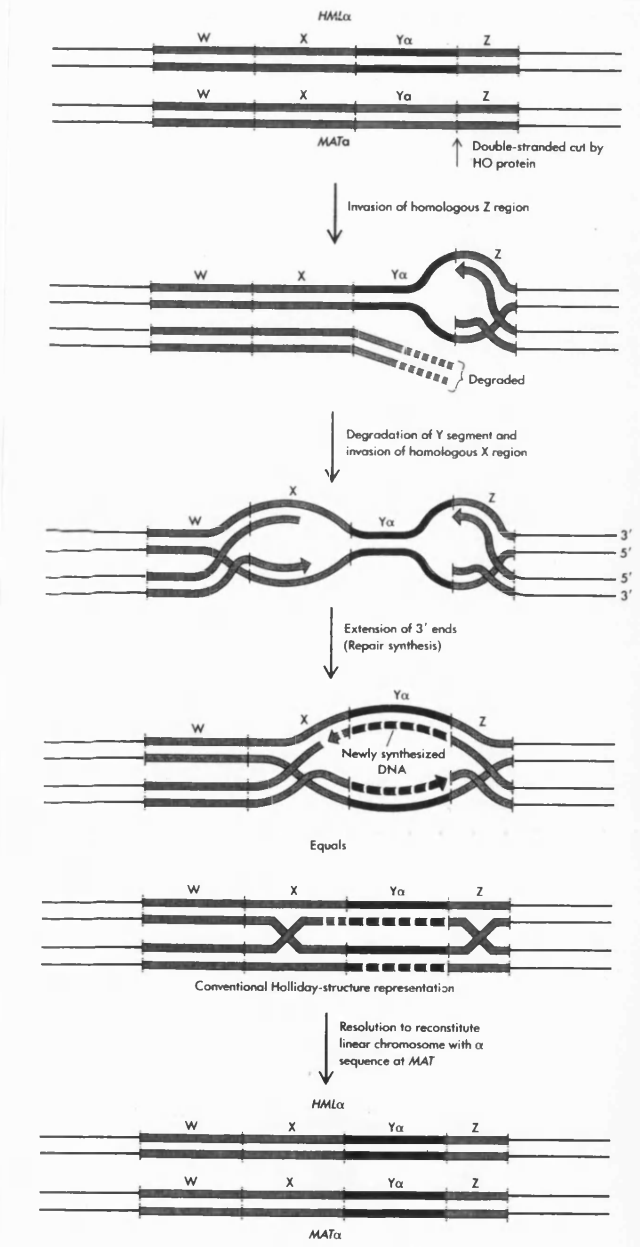


Figure 1.9 A basic model for *MAT* switching based on the double-strand break repair model of Szoztak *et al.* (1983). Described in main text. Reproduced from Watson *et al.* (1987).

Cassette switching (described with respect to Figure 1.9) is initiated by a double-strand break at the Y-Z1 junction by the HO endonuclease. The non-homologous Ya segment is 'chewed back' by exonuclease activity, leaving the free ends of the Z and X segments to invade the homologous regions of the *HMLα* gene, creating short heteroduplexes. The 3' extension during DNA repair replaces the missing Y segment with that of the invading *HM* gene. A symmetric heteroduplex is formed containing two Holliday junctions (Holliday, 1964) that must be resolved for the completion of switching. An unusual aspect of *MAT* switching is that it is very rarely (1%) accompanied by crossing-over (which would create a lethal deletion of a large chromosomal region), in contrast to gene conversion by double-strand gap repair, which is associated approximately 50% of the time with crossing-over (Klar and Strathern, 1984; Szostak *et al.*, 1983). In order to prevent isomerization (in this model) the Holliday junctions must be resolved in the same orientation, i.e. either the two noncrossed strands or the two crossed strands must be cut. This could theoretically occur via the action of a resolvase, specific for the mating-type switch reaction (Paques and Haber, 1999). Other mechanisms proposed to account for the low frequency of crossing-over include the action of topoisomerases, which might unwind both newly synthesised strands from their *HML* template strand, so that *HML* contains its original DNA and *MAT* carries the two newly synthesised regions, and there is no crossing-over (McGill *et al.*, 1989), or resolution of the two strands by a replication fork passing through the intermediate (Strathern *et al.*, 1982). However, this final proposal was disproved by Raveh *et al.* (1989), who demonstrated that heterothallic cells arrested in G1 with α-factor were able to successfully switch mating-type, when HO was expressed from a galactose inducible plasmid. Raveh *et al.* (1989) further demonstrated that the HO-cleaved *MAT* intermediate was stable over significant lengths of time, i.e. HO cleavage is not coupled to the subsequent gene conversion event, and that *rad52* mutants (defective in DNA damage repair) did not produce *MAT* DNA of the opposite mating-type, indicating that the block is prior to the gene duplication stage of the switch.

By using a galactose inducible HO endonuclease and physical monitoring techniques, the sequence of molecular events that follow the production of the double-strand break have been followed in considerable detail (reviewed by Haber, 1995). A key study by White and Haber (1990) identified several intermediate steps in the switching process (which was found to take 60 minutes from the appearance of

the double-strand break to the appearance of the switched product), thus: A long 3' single-stranded tail is produced in the Z region (and into unique sequences distal to *MAT*) by a slow (1 - 2 nucleotides per second) exonuclease, with no equivalent degradation of the proximal side of the double-strand break. The 3' end of *MAT*-Z is not significantly degraded. Once a 3' single-stranded tail has been produced, it can invade the donor locus and initiate repair of the double-strand break. An intermediate that can only be detected following strand invasion and DNA synthesis can first be detected 30 minutes before the completion of switching (this step is prohibited in *rad52* mutants). The removal of the original Y region, which must be excised before switching can be completed, only occurs as the completed product appears. The intermediates identified and the physical monitoring techniques used to do so, are summarized in Figure 1.10.

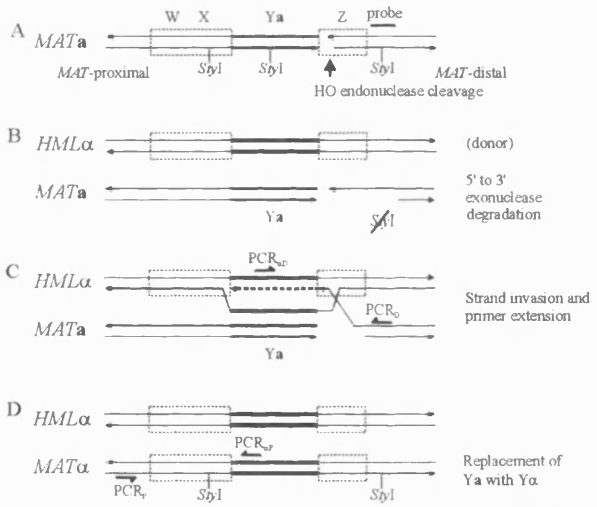


Figure 1.10 Molecular model of *MAT* switching based on the results of White and Haber (1990). HO endonuclease induces a double-strand cut within the *MAT* locus at the Y/Z junction. Exonucleolytic degradation to the right of the double-strand break creates a 3' single-stranded tail that invades the homologous sequences of the silent copy donor *HML*-Z region. Strand invasion and DNA extension can be monitored by a PCR assay using primers in *HML***α** and distal to *MAT*-Z (labelled $PCR_{\alpha D}$ and PCR_D). When strand extension reaches the X region of *HML***α**, the displaced strand anneals to the X sequence from *MAT***a** and serves as a template for the replacement of the Y**a** sequence with Y**α** sequence. Formation of the new *MAT*-X/Y**a** junction can be monitored using PCR primers located in the unique *MAT* proximal sequence (PCR_P) and in the Y**α** sequence ($PCR_{\alpha P}$). The kinetics of strand degradation can be followed by Southern hybridization using a probe that hybridizes to *MAT* and to the adjacent distal fragment. As the single-stranded tail is produced, a larger restriction fragment, refractory to *Sty*I cleavage, at one or more distal sites, is produced. Reproduced from Sugawara *et al.* (1995).

The sequence of events (above) contrast to those described for the simple double-strand gap repair model for switching (Figure 1.9). *MAT* switching is driven by recombination events initiated by *MAT*-Z, with both the removal of the original Y region and the joining of the newly copied sequences to *MAT*-X being late events. Other studies have revealed further insights into the switching reaction, thus: The region replaced during *MAT* switching is substantially longer than the Y region itself. McGill *et al.* (1989) used artificial restriction sites inserted at different places in the X and Z regions to show that replacement of the Y segment often extends into the flanking homologous regions. They further showed that there was no reciprocal transfer of markers from *MAT* to the donor. This observation was supported by studies of a single base pair mutation only 8 b.p. from the 3' of the HO cut, in the Z region (Ray *et al.*, 1991). In the absence of mismatch repair (Section 1.3), this mutation was most often retained during switching (thus confirming physical studies showing that there was almost no 3' to 5' removal of the 3'-ended tail). Usually, only one of the two daughter cells carried the mutation. Kinetic analysis (Haber *et al.*, 1993) further demonstrated that, in repair-proficient cells, mismatch correction occurred very rapidly (as quickly as the PCR-amplified intermediate could be detected), suggesting that correction occurred soon after the strand invaded the donor locus. Moreover, the heteroduplex DNA was corrected in a highly biased way, such that mutant sequence in the invading Z DNA was corrected to the genotype of the donor. Sugawara *et al.* (1995) used physical monitoring of DNA to show that *MAT* switching is completely blocked at an early step in recombination in strains deleted for the DNA repair genes *RAD51, RAD52, RAD54, RAD55* or *RAD57* (initially identified because of their role in repairing X-ray induced damage). When the donor sequence was simultaneously not silenced and located on a plasmid, only the *RAD52* gene was required. *RAD51, RAD54, RAD55* and *RAD57* are still required when the same transcribed donor is on the chromosome. One puzzling feature of *MAT* switching is that ends of HO-cleaved *MAT* DNA are able to invade the same region of *HML* or *HMR* where HO cannot cut. It was shown that *RAD51, RAD54, RAD55* and *RAD57* are all required to facilitate strand invasion into otherwise inaccessible regions of DNA.

Physical monitoring of *MAT* switching has made it possible to identify the early steps in the process, but exactly what happens in some later steps remains unclear. For example, how are the original Y sequences, which are not homologous to

the incoming Y sequences, removed? The investigation by White and Haber (1990), had suggested that neither strand of Y**a** was degraded until just before switching is completed, whereupon both strands disappeared (Figure 1.10). However, later analysis suggests that, as on the distal side of the double-strand break, one strand is degraded 5' to 3', producing a long single-stranded tail (Wu *et al.*, 1997). Physical analysis of the rate of 5' to 3' degradation has implicated the proteins, Rad50p, Mre11p and Xrs2p, in this process (Ivanov *et al.*, 1994; Sugawara and Haber, 1992; Tsubouchi and Ogawa, 1998). It is suggested by Haber (1998b) that removal of the single-stranded tail occurs by endonucleolytic cleavage at its junction with paired, homologous sequences. A process that is dependant on two proteins from the nucleotide excision repair pathway (Rad1p/Rad10p) and two proteins from the mismatch repair pathway (Msh2p/Msh3p) [Fishman-Lobell and Haber, 1992].

Physical analysis has also made it possible to analyse conditional lethal mutants to ask what DNA replication enzymes are required for *MAT* switching. Many of the proteins required for origin dependent DNA replication are required for *MAT* switching, which include the single-strand binding protein RPA, the clamp protein PCNA (an elongation factor), and possibly all three major DNA polymerases (Haber, 1998b). The most striking result is that *MAT* switching appears to involve both leading and lagging strand synthesis, mutations in primase and polymerase α drastically impair the completion of switching.

The absence of crossing-over, the lack of transfer of genetic information from *MAT* to the donors during recombination, and the involvement of lagging strand synthesis are consistent with a new family of double-strand break repair models (reviewed by Paques and Haber, 1999), known collectively as synthesis-dependent strand annealing (Haber, 1998b). One version of this mechanism that accounts for the involvement of lagging strand synthesis is illustrated in Figure 1.11. In this mechanism, the strand invasion sets up a replication fork in which there is both leading and lagging strand synthesis. Unlike normal origin-dependent DNA replication, however, branch migration is postulated to displace both newly synthesised strands so that they are both inherited by the recipient (*MAT*) locus, leaving the donor completely unchanged. This mechanism does not lead to the formation of a pair of Holliday junctions that can be resolved by crossing-over, so that crossing-over would rarely accompany *MAT* switching.

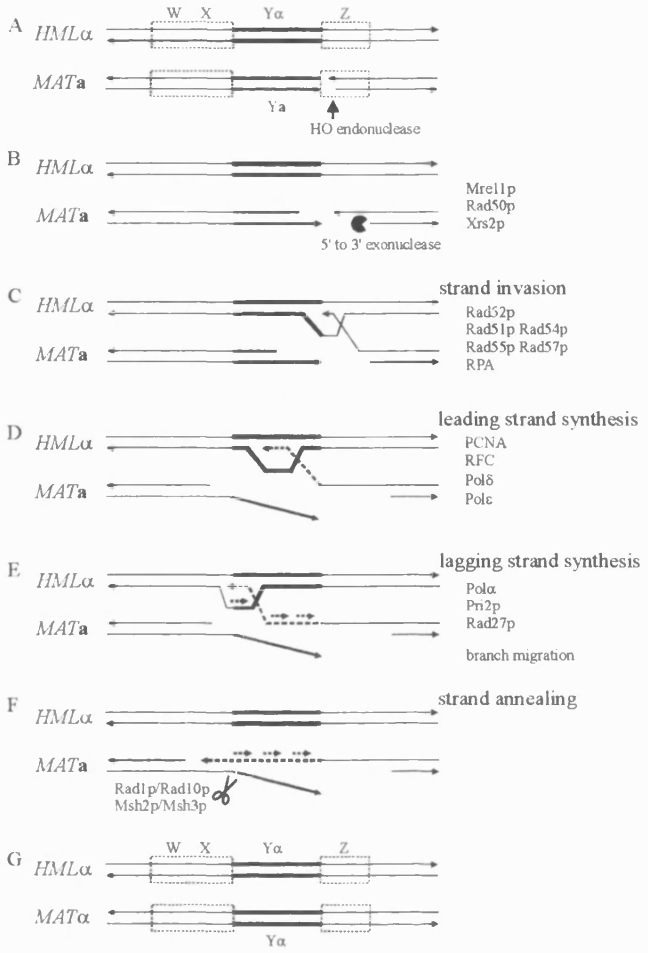


Figure 1.11 Synthesis-dependent strand annealing model for *MAT* switching. Proteins involved at different stages in the process are shown. In this mechanism, *RAD52*-dependent strand invasion initiates new DNA synthesis, requiring RFC (replication factor C), PCNA, and either Polδ or Polε. The newly synthesised strand is displaced and eventually anneals with the second broken end, without the formation of a Holliday junction that would permit crossing-over to accompany *MAT* switching. A second strand may be synthesised by lagging strand replication. Removal of the original Y region is excised by a 'flap' endonuclease including Rad1p/Rad10p and Msh2p/Msh3p. Reproduced from Haber, 1998b).

Donor preference

As previously discussed, the mating-type switching reaction is a directed process, one that is independent of the sequences at the silent donor loci. Thus, *MAT***a** cells preferentially recombine with *HML*, whilst *MAT*α cells chose to recombine with *HMR*, ensuring that recombinational repair of the double-strand break will lead to a switch to the opposite mating-type. The underlying mechanism governing donor selection have been investigated, principal aspects of which are described below.

Wu and Haber (1995) investigated donor preference in *MAT***a** cells. By deleting *HML* and inserting a donor at other chromosome III locations, it was found that *MAT***a** activated a large (>40 k.b.p.) region near the left end of the chromosome, such that a donor placed at several sites within this domain was strongly preferred over *HMR*. When inserted outside of this domain, the donor was used equally with *HMR*. However, *MAT***a** donor preference for *HML* was found to be abolished by the expression of the negative regulator, *MAT*α2 (although *HML* regained its preferred status when the donor was unsilenced). It was also shown that mating-type dependent activation of the left end of the chromosome applied to other recombination events. Spontaneous recombination between two *leu2* alleles was found to be 20 - 30 times higher in *MAT***a** than in *MAT*α when one of the *leu2* alleles was inserted instead of *HML*.

In order to locate *cis*-acting elements responsible for the preference of *MAT***a** cells for *HML*, Wu and Haber (1996), created a series of terminal truncation (by including a telomere sequence) or internal deletions. A 2.5 k.b.p. deletion located 17 k.b.p. proximal to *HML* was found to completely reverses donor preference, so that a *MAT***a** cell used *HML* only 10% of the time instead of 90%. By re-inserting subfragments of this deleted region, a 700 b.p. 'recombination enhancer' (RE) sequence that controls *MAT***a** and *MAT*α donor preference was defined. The recombination enhancer was found not only to activate *HML* at its normal location but also to activate either *HML* or *HMR* sequences placed at several different locations along the entire left arm of chromosome III. In each case, *HMR* became the strongly preferred donor, if the recombination enhancer was deleted, as in *MAT*α cells. Thus the left arm of chromosome III was shown to be recombinationally inactive in *MAT*α and somehow activated in *MAT***a**. The 'cold' state in *MAT*α, whether the recombination enhancer is present or not, also being the default state in *MAT***a** cells deleted for the recombination enhancer. The recombination enhancer acting to reverse the unusual inaccessibility of donors on the left arm of chromosome III. The role of the recombination enhancer is summarized in Figure 1.12.

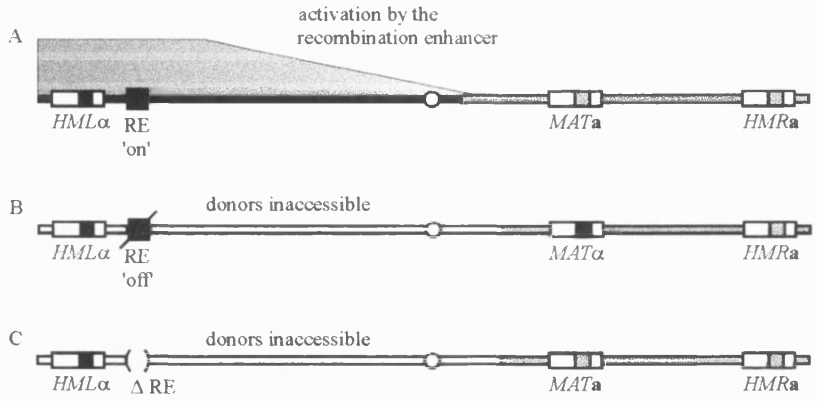


Figure 1.12 Donor preference is regulated by the recombination enhancer RE. **A**. In *MAT***a** cells, a donor placed anywhere along the left arm of chromosome III is used preferentially in competition with *HMR*. **B**. In *MAT*α cells, the RE is turned off and the entire left arm and part of the right arm become cold, allowing *HMR* to be the preferred donor. **C**. When the RE is deleted in a *MAT***a** cell, the left arm becomes inaccessible and *HMR* becomes the preferred donor. Reproduced from Haber (1988a).

By comparing the recombination enhancer sequences of *S. cerevisiae* and *S. carlsbergensis* (which is functional in *S. cerevisiae*), the recombination enhancer was narrowed down to 244 b.p., and found to contain four well-conserved sub-domains (Wu *et al.*, 1998). Deletion of three of the regions, A, C or D, abolished recombination enhancer activity. Region C contained a highly conserved binding site for the Matα2p-Mcm1p co-repressor (the repressor of **a**-specific genes in *MAT*α cells). The Matα2p-Mcm1p co-repressor has been shown to play a key role in repressing the activity of the recombination enhancer in *MAT*α cells, whilst Mcm1p (transcriptional activator of **a**-specific genes) has been shown to play a definitive role in activation of the recombination enhancer in *MAT***a** cells (Szeto *et al.*,1997; Wu *et al.*, 1998).

The genetic studies of how the recombination enhancer is regulated (above) are supported by analysis of the chromatin structure of the recombination enhancer. In *MAT*α cells, the recombination enhancer is covered by a very highly positioned set of nucleosomes on either side of the (occupied) Matα2p-Mcm1p binding site (Weiss and Simpson, 1997). The ordered array extends across the entire 2.5 k.b.p. interval containing the recombination enhancer, and ends at two flanking open reading frames. In *MAT***a** cells, the recombination enhancer exhibits several distinctive footprints indicative of protein binding and a region of closely spaced DNaseI hypersensitive sites. In *MAT*α cells, when mutations of the MATα2p binding sites of the Matα2p-Mcm1p binding domain markedly increase the use of *HML* as a donor, the

chromatin structure of the recombination enhancer region becomes indistinguishable from that seen in *MAT***a** cells (Wu and Haber, 1995).

It is proposed by Haber (1998a) that the recombination enhancer might function by changing the nuclear localisation or the higher-order folding of the entire left arm of chromosome III to make it more 'flexible' in locating and pairing with the recipient site in *MAT***a** cells. In this view, the chromosome arm would be sequestered or immobilized (perhaps by being bound to the nuclear envelope) in such a way that *HML* is unavailable in *MAT*α cells even though the chromatin structure at *HML* itself remains unchanged.

1.2 *IN VITRO* DIRECTED MOLECULAR EVOLUTION

The stunning array of features and functions exhibited by proteins and enzymes makes them valuable and attractive resources for use in medical applications and industrial processes. However, proteins isolated from natural sources, when recruited for new tasks, e.g. as catalysts in industrial processes or as additives in laundry detergent, are often not well suited. Desirable features might include higher thermal stability, the ability to function in organic solvents, in extremes of pH or under oxidative conditions, as well as high catalytic activity to (or the ability to recognize) non-natural substrates (Arnold, 1999).

The possibility of designing a new protein from first principles to perform a specific task is remote. Despite extensive investigation, our understanding of the mechanisms of protein folding and function in relation to the structure is in its infancy (Harayama, 1998). To this end, protein engineers have attempted to modify the properties of existing proteins, a process made possible by two techniques: site-directed mutagenesis (Reidhaar-Olson *et al.*, 1991) and computer-assisted modelling (Clealand, 1996). However, this approach is only applicable to protein families in which the three-dimensional structure of at least one member protein has been resolved. Furthermore, many attempts to alter the properties of enzymes by 'rational design' have failed, because the introduced amino acid substitutions exerted unexpected influences on the structure and function of the target enzymes.

A more powerful method of altering a protein toward a desired goal is directed molecular evolution. Here no structural information is required and there is no need to identify amino acid residues that are important to function. Instead the protein or enzyme is subjected to repeated operation of three processes: selection, amplification and mutation, fundamentally Darwinian evolution, until the desired protein is obtained.

One way this can be achieved is by using error-prone polymerase chain reaction (PCR), which allows the introduction of random mutations, followed by selection/screening[Ψ] for desired variants. Chen and Arnold (1993) used this approach to engineer the protease subtilisin E to function in high concentrations of the polar organic solvent dimethyl formamide (DMF), in which it catalyses peptide bond formation. Three sequential rounds of mutagenesis and screening (starting with a

variant containing four effective amino acid substitutions) yielded a variant containing ten amino acid substitutions that exhibited more than 150 times the activity of wild-type subtilisin in 60% DMF. Two additional rounds of mutagenesis and selection (You and Arnold, 1994) allowed the isolation of a mutant enzyme containing three further amino acid substitutions that was 471 times more active in 60% DMF than the wild-type enzyme. The 13 amino acid substitutions resulted from 27 DNA base substitutions distributed across the 800 b.p. portion of the gene sequenced.

A critical point of this method, as discussed by Moore and Arnold (1996), is that the mutation frequency must be carefully tuned: generally beneficial mutations are rare, while deleterious mutations are common; because the combination of a beneficial mutation and a deleterious mutation may result in the formation of an inactive enzyme, the mutation frequency should not be high if the maximum number of improved enzymes are to be obtained. The mutation frequency should also not be too low, otherwise the wild-type sequence would predominate in the population.

The problem of deleterious mutations accumulating with successive rounds of mutagenesis was overcome by the 'DNA shuffling' technique devised by Willem Stemmer (Stemmer, 1994a; Stemmer 1994b). The technique, described briefly earlier, is combinatorial in its approach: A gene (or pool of mutated genes) is randomly fragmented with DNaseI. These fragments (30 b.p. - 300 b.p. in size) are then reassembled into a full-length gene by repeated cycles of annealing in the presence of DNA polymerase. The fragments prime with each other based on homology, and recombination occurs when fragments from one copy of a gene prime on another copy, causing a template switch. The method produces a point mutagenesis rate of 0.7%, similar to error-prone PCR. As in random mutagenesis, mutants are screened, and improved variants used as templates for recursive rounds of shuffling, assembly and selection (Figure 1.13).

[ψ] Selection strategies are, in principle, much more powerful than screens for analysing and improving molecular functions, because huge number of variants can be screened simultaneously. In a screen, variants are assayed individually for a given property to identify the best. In a selection, a pool containing all the variants is subjected to a selection step that yields the 'best' molecule directly (Clarkson and Wells, 1994). Most of the experimental effort of directed evolution is devoted to devising, validating and implementing a suitable screen, this is due to the first law of random mutagenesis "you only get what you screen for" (Arnold, 1998a).

The technique is more powerful as it introduces an element missing in random mutagenesis, but prevalent in natural evolution: sex (correspondingly the technique has also been called 'sexual PCR' [Smith, 1994]; although the term 'chain reaction' is a misnomer, as there is no autocatalytic amplification). Recombination between different DNA lineages allows beneficial mutations to come together and the elimination of deleterious ones. DNA shuffling also permits a 'backcross' (with an excess of the wild-type strand) to be made, allowing selection against any mutations that are not important for function.

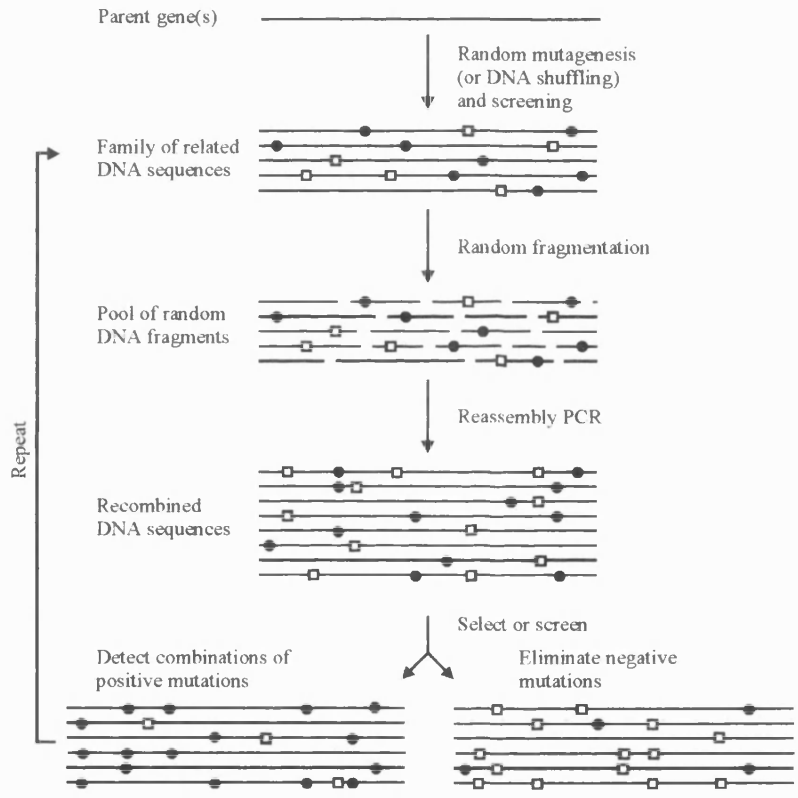


Figure 1.13 *In vitro* recombination by DNA shuffling. Random fragments of a single parent gene or pool of genes containing different mutations are reassembled in a PCR reaction, during which they are allowed to act as primers for one another. The fragments elongate, producing a family of DNA sequences containing different combinations of the mutations contained in the parent pool. Beneficial mutations are accumulated and deleterious mutations eliminated during recursive cycles of shuffling and screening or selection. Reproduced from Arnold (1997).

In his study, Stemmer started with a β-lactamase gene, TEM-1, with very low activity against the antibiotic cefotaxime. The minimum inhibitory concentration for *Escherichia coli* expressing the gene from a plasmid was 0.02 $\mu g ml^{-1}$. The gene was

subjected to DNA shuffling, the PCR product spliced back into the plasmid vector, the spliced vector transformed into *E. coli*, and the transformed bacteria selected on various concentrations of cefotaxime. Several hundred colonies showing the highest levels of resistance provided the starting DNA for another cycle of the same four steps. After three cycles of mutagenesis, recombination and selection, the minimum inhibitory concentration for the most resistant clones was 640 μgml^{-1}, a 32,000-fold increase. Following two backcross cycles, with a 40-fold excess of wild-type, the mutant gene conferring high resistance contained six amino-acid changes, plus a promoter change. A control experiment using error-prone PCR but no shuffling resulted in a minimum inhibitory concentration, after three selection cycles, of only 0.32 μgml^{-1}, a 20-fold increase.

Diversity in the DNA shuffling reaction originates from random point mutations, a result of the polymerase reaction. Most of these mutations are (as discussed above) deleterious or neutral. Naturally occurring homologous sequences, however, are pre-enriched for 'functional diversity', because deleterious variants have been selected against over billions of years of evolution. Crameri *et al.* (1998) showed that shuffling gene families (Figure 1.14) was an effective way to accelerate the process of directed evolution: Four cephalosporinase genes - 58 to 82% identical at DNA sequence level - from four microbial species were evolved separately, and as a mixed pool, for increased resistance against the antibiotic moxalactam. Following a single round of shuffling, clones originating from the four single gene libraries showed up to eightfold increases in moxalactam resistance, as compared to those expressing the wild-type genes. In contrast, the best clone originating from the gene family library showed a 540-fold increase (0.38 - 200 μgml^{-1}) to the 'best' wild-type genes, and a 270-fold increase (0.75 - 200 μgml^{-1}) to the 'worst' wild-type genes. Thus 'family shuffling' accelerated the rate of functional enzyme improvement 34- to 68-fold in a single cycle.

The utility of family DNA shuffling was demonstrated by Chang *et al.* (1999). Human alpha interferons (Hu-IFN-αs) are encoded by a family of over 20 tandemly duplicated nonallelic genes that share 85 - 98% sequence identity at the amino acid level. These proteins have potent antiviral and antiproliferative activities that have clinical value as anticancer and antiviral therapeutics, e.g. Hu-IFN-α2a is marketed as a $600 million/year drug by Schering-Plough of Berlin (Coghlan, 1998). Following two rounds of DNA shuffling, in which all of the Hu-IFN-α genes, including

pseudogenes, were shuffled, the most active clone was improved 285,000-fold relative to Hu-IFN-α2a, as measured by its ability to protect cultured cells against a mouse virus.

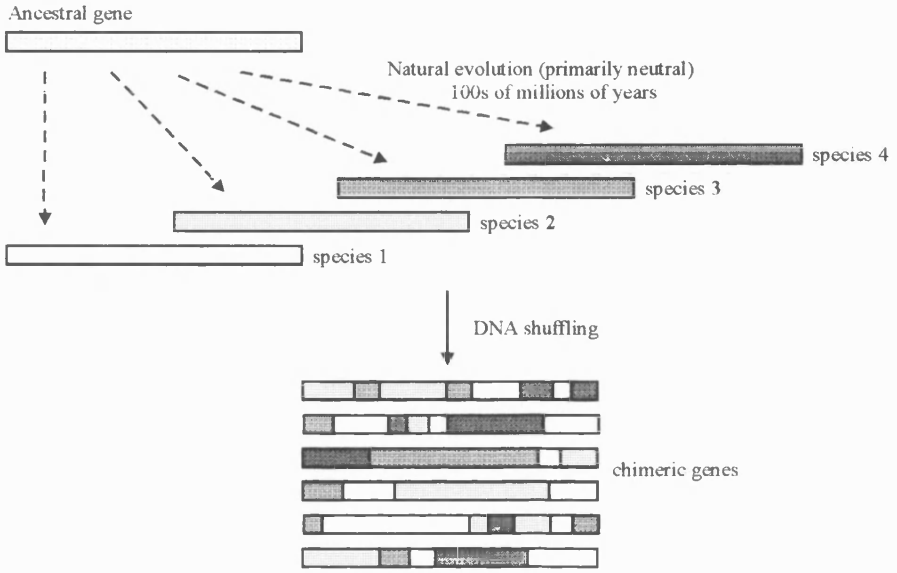


Figure 1.14 Family DNA shuffling. Homologous proteins are descended from a common ancestral protein and share its overall three-dimensional structure. Recombining homologous genes creates chimeras, some fraction of which should also fold into that structure. Such family shuffled libraries could be rich in novel function. Reproduced from Arnold (1998b).

The applicability of DNA shuffling to the evolution of different characteristics has been demonstrated, for example: Following three rounds of DNA shuffling, the fluorescence of the green fluorescent protein (GFP) of the jellyfish was improved 45-fold relative to that of the initial GFP (Crameri *et al.*, 1996). Substrate specificity of the *E. coli lacZ* gene was changed, following seven rounds of shuffling, from a β-galactosidase into an efficient, 1,000-fold improved, β-fucosidase (Zhang *et al.*, 1997). An arsenate detoxification pathway, conferred by a plasmid borne operon, was improved 40-fold following three rounds of shuffling, enabling *E. coli* cells to grow on 0.5 M arsenate. (Crameri *et al.*, 1997). Using a combination of DNA shuffling, random mutagenesis and recombination via restriction sites, Giver *et al.* (1998) increased the thermostability of an esterase by >14°C, whilst maintaining catalytic activity at low temperatures.

DNA shuffling has also been applied to another important protein engineering technology; phage display. Devised by McCafferty *et al.* (1990), phage display

(described briefly here) is a method for making antibodies *in vitro* by mimicking the immune system. The strategy relies on the surface display by filamentous bacteriophage of a combinatorial library of antibody fragments. There are many versions of the phage-display technique (Winter *et al.*, 1994), a typical one (as devised by Marks *et al.*, 1991), is as follows: A diverse library of immunoglobulin heavy (V_H) and light (V_K and V_λ) chain variable (V) genes (which code the antigen-binding site of antibodies) are prepared from peripheral-blood cells by PCR amplification, using universal primers. Genes encoding single chain Fv fragments are made by randomly combining heavy and light chain V-genes using PCR, and the combinatorial library ($>10^7$ members) cloned for display on the surface of a phage. Rare phage that bind to an antigen of interest (with good affinity) are selected by rounds of growth and panning with antigen.

Combinatorial infection (Waterhouse *et al.*, 1993) has been used to generate large phage libraries (up to 10^{12} members) allowing high-affinity antibodies to be selected directly. However, rather than prepare a large library of V_L and V_H genes, Crameri *et al.* (1996) showed that a strategy based on DNA shuffling might be used to evolve V_L and V_H genes. V_L and V_H consist of three hypervariable regions (complementarity-determining regions, or CDRs) flanked by four relatively conserved regions (framework regions, or FRs). DNA shuffling was designed so that mutated CDRs were randomly recombined at FRs to create new combinations. In this way high-specificity antibodies recognising a specific human receptor were obtained, following eight cycles of shuffling and selection.

In vitro recombination by staggered extension

An alternative strategy to that of DNA shuffling for recombining mutated sequences was devised by Zhao *et al.* (1998). The staggered extension process (StEP) does not require fragmentation of the template sequences, instead, recombination is promoted between templates in a modified PCR type reaction. StEP consists of priming template sequences followed by repeated cycles of denaturation and extremely abbreviated annealing/polymerase-catalysed extension (Figure 1.15). In each cycle the growing fragment can anneal to different templates based on sequence complementarity and extend further to create 'recombination cassettes'. Due to the template switching the growing polynucleotides contain sequence information from different parental genes. StEP is continued until full-length genes are formed.

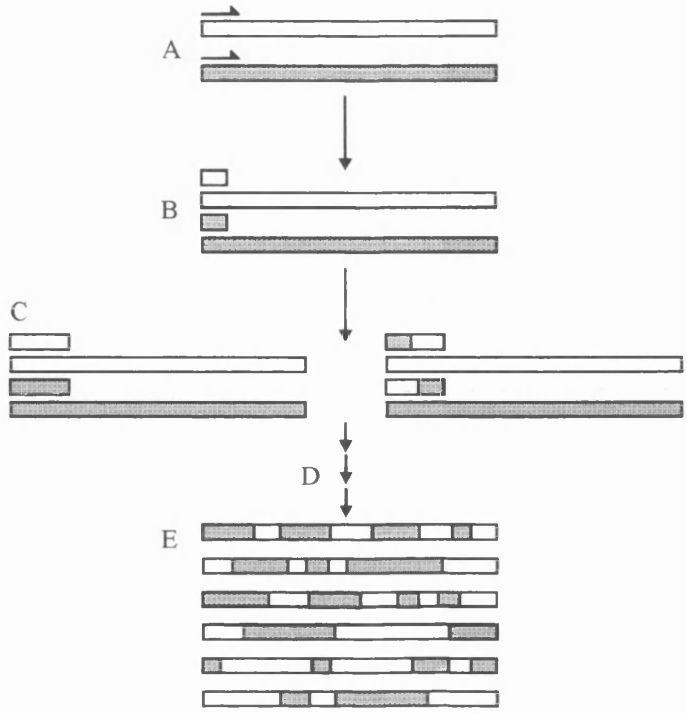


Figure 1.15 StEP recombination. Only one primer and single strands from two parent genes (templates) are shown. **A**. Denatured template genes are primed with one defined primer. **B**. Short fragments are produced by brief polymerase-catalyzed primer extension. **C**. Through another cycle of StEP, fragments randomly prime the templates (template switching) and extend further. **D**. Process is repeated until full-length genes are produced. **E**. Full-length genes are purified and (optionally) amplified in a PCR reaction with external primers. Reproduced from Zhao *et al.* (1998).

Following a round of mutagenic PCR and selection, StEP represents a convenient method to recombine variant (improved) sequences. Unlike DNA shuffling, the StEP reaction itself is not mutagenic, point mutations arise at the same frequency as observed for *Taq* polymerase, ~0.1%. In their study, Zhao *et al.* (1998) showed that the staggered extension process could efficiently recombine mutations as close as 35 b.p., a similar efficiency as that obtained for DNA shuffling. StEP recombination of five thermostable subtilisin E variants (isolated following mutagenic PCR and selection) yielded a second-generation subtilisin E variant that possessed a half-life 50-times that of wild-type subtilisin E, at 65°C.

Molecular evolution using a combination of mutagenic PCR and StEP has been successfully applied. Examples (among others) include: Zhao and Arnold (1999), who used StEP to probe the relationship between enzyme structure and thermostability; Lin *et al.* (1999), who evolved horseradish peroxidase so that it was

functionally expressed in *E. coli*, and Joo *et al.* (1999), who evolved the monooxygenase cytochrome-P450 so that it functioned in the absence of its cofactors.

Other strategies

The number of strategies devised for performing directed evolution and for the engineering of novel proteins continues to grow. Some of these strategies are introduced below.

Random-priming recombination (RPR), is a recombination strategy similar to StEP, except that random primers (hexamers) are used instead of specific primers (Shao *et al.*, 1998). Frequent template switching occurs as the primers extend, and prime each other (in a similar manner to fragment reassembly in DNA shuffling) resulting in full-length shuffled genes. The number of cross-overs per gene for RPR ranges from one to six, compared with one to four for StEP.

RACHIT (random chimeragenesis on transient templates) is a technique that aims to maximise the frequency of crossing-over (Pelletier, 2001). A pool of homologous genes are fragmented by digestion with DNaseI. However, unlike DNA shuffling, the DNAs used are single stranded. Full-length genes are reassembled by hybridisation (in the absence of polymerase) to a single stranded scaffold, designed to be homologous, but not identical, to any of the fragmented parental genes. Overlaps in the hybridised fragments are removed by digestion, and the gaps between fragments filled in by polymerase. The fragments are then ligated, yielding a pool of full-length, diversified single strands hybridised to the scaffold, with an average of 14 cross-overs per gene (compared with one to four for DNA shuffling). A unique feature of the technique is that the scaffold can be digested (due to incorporation of uracil), precluding its subsequent amplification, thus reducing the background of unshuffled genes.

In random elongation mutagenesis (REM), randomized oligonucleotides are incorporated ahead of the stop codon, exploiting the stabilising effect (in certain proteins) of terminal arms of amino acids that fold across the susceptible part of the structure (Matsuura *et al.*, 1999). This technique allows the 'fitness landscape' of the protein being evolved to be expanded, overcoming one of the limitations of DNA shuffling and random mutagenesis, where large jumps in sequence space cannot be made.

Another limitation of DNA shuffling is that the total combinatorial space cannot be exploited, as crossing-over only occurs in regions of relatively high homology. ITCHY (incremental truncation for the creation of hybrid enzymes) is a method that allows access to more diverse sequence space (Ostermeier *et al.*, 1999b). The strategy relies on a slow (≤10 bases/min) exonuclease digestion to create a random-length 5' fragment library of the gene encoding protein A and a random-length 3' fragment library of the gene encoding protein B. The truncated fragments are then ligated, forming a library of gene fusions, in which cross-overs occur in a manner independent of sequence homology (Figure 1.16).

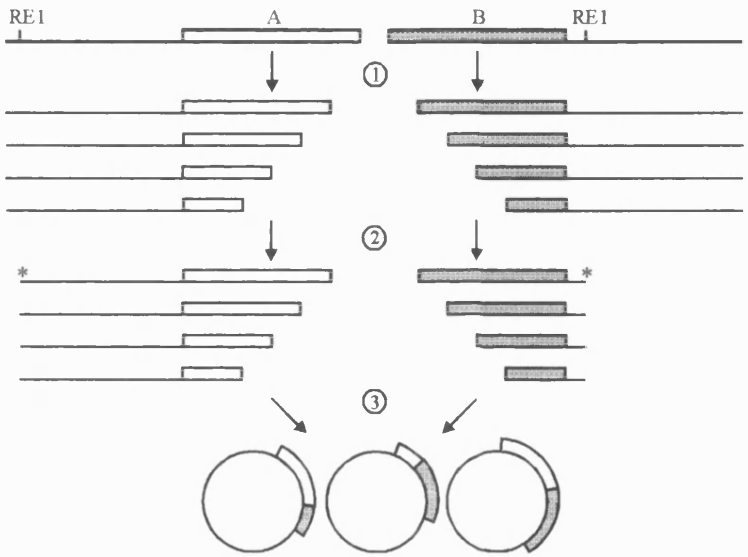


Figure 1.16 The ITCHY procedure. Genes A and B are cloned into two unique expression vectors and linearized. 1. An incrementally truncated library (ITL) is generated by slow exonuclease digestion. 2. The linear ITLs are digested with a restriction enzyme RE1. 3. The fragments containing the ITL of B are ligated into the vector containing the ITL of A, this proceeds via sticky end ligation at the asterix and a blunt end ligation between the truncated genes. Reproduced from Ostermeier *et al.* (1999a).

The ITCHY strategy, however, suffers from the drawback, namely that gene length is not conserved, as random-length fragments are fused to other random-length fragments. Only a tiny fraction of the cross-overs connect the two parent genes at sites where the sequences align and are structurally related. An alternative (and possibly simpler) strategy to ITCHY, which retains proper sequence alignment, has recently been reported by Sieber *et al.* (2001). In SHIPREC (sequence homology-independent protein recombination) two parent genes are fused to form a

gene dimer, the gene dimer is then digested with DNaseI to form an ensemble of random-length fragments. Fragments corresponding in length to either of the parent sequences are isolated (ensuring that the two amino acids that meet at the cross-over are in similar positions in the parent structure), blunt ended, and the remaining ends fused by circularization. The gene that was at the 5' position in the dimer will now be at the 3' position and will donate the C terminus of the hybrid protein. The completed library is generated by linearization of the circular DNA pieces by restriction digestion (Figure 1.17).

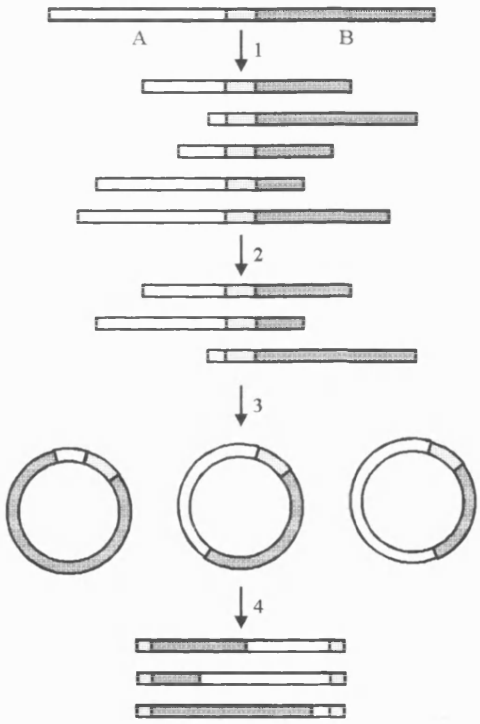


Figure 1.17 The SHIPREC procedure. A gene dimer is constructed that consists of (from 5' to 3') the gene of protein A, a linker sequence containing restriction sites, the gene for protein B. 1. The dimer is fragmented by digestion with DNaseI and treated to produce blunt ends. 2. Fragments of the length of a single gene (plus the length of the linker sequence) are separated from the pool. 3. Single-gene length fragments are circularized by intramolecular blunt-end ligation. 4. Linearization of the circular DNA pieces by restriction digestion in the linker region creates a library of chimeric genes with members having an N-terminal from protein B and a C-terminal part from protein A, the cross-overs being distributed over the entire length of the gene. Finally the chimeric genes are cloned into an expression vector. Reproduced form Sieber *et al.* (2001).

Finally, another interesting approach for the creation of novel enzymes is exon shuffling, reviewed by Kolkman and Stemmer (2001). Mosaic proteins, e.g. proteins

involved in blood-clotting (Figure 1.18), are composed of a number of discrete domains. The individual domains are often involved in specific functions that contribute to the overall activity of the protein. The domains found in mosaic proteins are evolutionarily mobile and occur in otherwise unrelated proteins. The mobile domains are characterised by their ability to fold independently, which is essential because it prevents misfolding when they are inserted into a new protein environment. Further, in many mosaic proteins each domain is encoded by one (or a combination) of exons. Rearrangements of exons, by recombination in the intervening intron sequences, was first proposed by Gilbert (1978), as a way in which genes for novel proteins arise in the eukaryotic genome. In exon shuffling this process is copied *in vitro*, with the aim of producing novel proteins, or improved variants of existing ones. More than 60 different mobile domains have been identified (Kolkman and Stemmer, 2001) which can be recombined, using PCR-based techniques, in a variety of different formats (Table 1.1) to produce libraries of exon shuffled genes, which can be screened for desirable functions or properties.

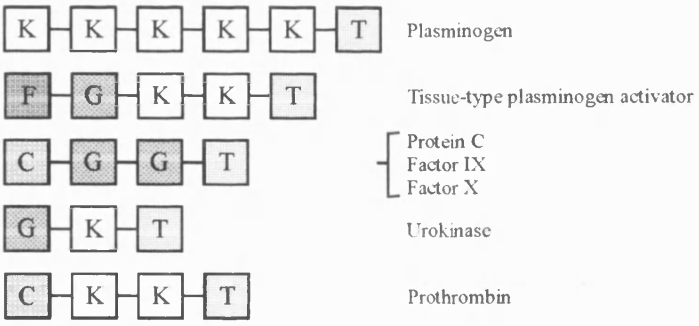


Figure 1.18 Exchange of modules by exon shuffling in proteases involved in fibrinolysis and blood coagulation. Plasminogen, tissue plasminogen activator, protein C, factor IX, factor X, urokinase and prothrombin consist of a trypsin-like serine-protease domain (T) and other distinct structural functional domains (K, kringle-domains; F, finger domains; G, epidermal-growth-factor-related domains; C, vitamin-K-dependent calcium-binding domains). Each domain can fold independently and is functionally autonomous. The structures of these proteases suggests that they were assembled from modules that had been encoded independently in discrete DNA fragments (exons). Figure and legend reproduced from Harayama (1998).

Examples of library formats based on exon shuffling

Orthologous exon shuffling	Substitution of exons with equivalent exons from the same gene of different species
Paralogous exon shuffling	Substitution of exons with homologous exons from different genes of the same species
Orthologous domain shuffling	Substitution of domains with equivalent domains encoded by the same gene of different species
Paralogous domain shuffling	Substitution of domains with homologs encoded by different genes of the same species
Domain number variation	Deletion duplication or insertion of protein domains
Functional homolog shuffling	Substitution of domains with non-homologous domains that are functionally related
De novo protein assembly	Assembly of multiple, independent domains into novel proteins

Table 1.1 Reproduced from Kolkman and Stemmer (2001).

1.3 EXPERIMENTAL STRATEGY

The synthetic donor/acceptor gene locus to be constructed is illustrated in Figure 1.19. The 'donor' and 'acceptor' genes (described below) will be integrated into the yeast genome (chromosome III). Initially, for simplicity, only one 'donor' (hereafter donor) gene will be present. To prevent donor gene expression it will be flanked by the *E* and *I* silencer elements, which ordinarily repress the mating-type genes (Section 1.1.1), this will be achieved by integrating the donor gene at the *HML* locus. The 'acceptor' (hereafter acceptor) gene will be integrated at the *LEU2* locus. Immediately following the termination codons of both the donor and acceptor genes is a rare restriction site (cut by the yeast mitochondrial endonuclease, I-*Sce*I [described below]; this will be maintained separately on a plasmid, under the control of the inducible promoter, *GAL1*). On induction it is intended that the endonuclease makes a double-strand break in the acceptor gene (as for the mating-type genes at *HML*, the silencer elements, as well as repressing the donor gene, will also protect the restriction site from being cut). The double-stranded break will then be repaired using the donor gene as template (the mode of this repair reaction is discussed below).

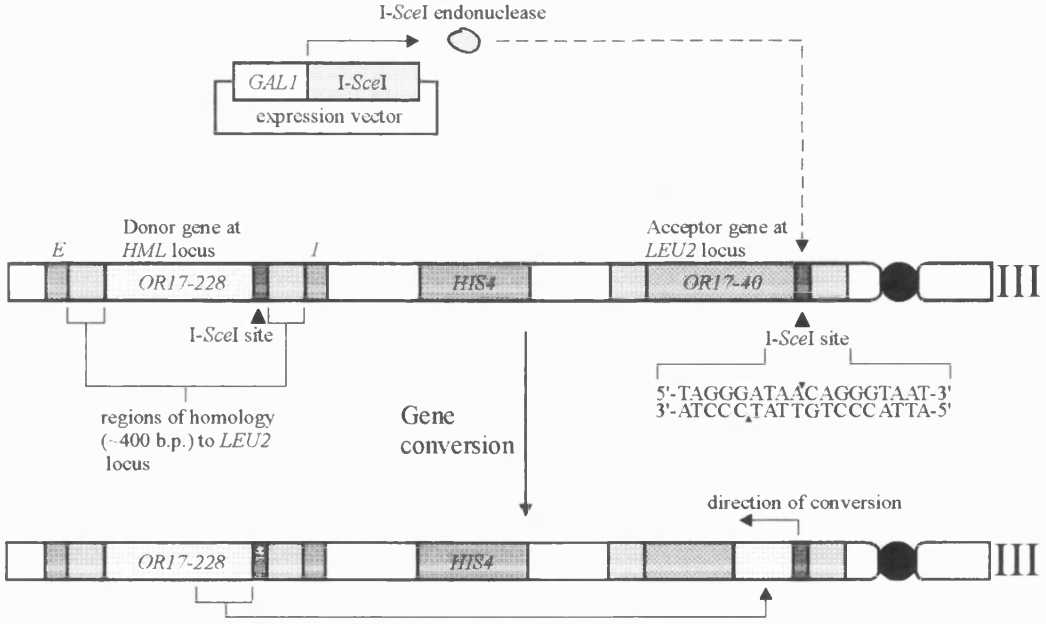


Figure 1.19 The synthetic donor/acceptor locus, described in main text (Figure is not to scale).

Similar experiments to that proposed have been successfully carried out with identical donor and acceptor genes (Ray *et al.*, 1988; Nickoloff *et al.*, 1989). In these experiments, deigned to investigated the association of intra-chromosomal crossing-over with gene conversion, a double-strand break was imparted *in vivo* (by the *HO* endonuclease) to one copy of a directly repeated gene. The proposed experiment differs from these previous experiments in that the donor and acceptor genes are homologous, not identical (although spontaneous recombination between homologous genes has been investigated, e.g. Bailis and Rothstein (1990), Datta *et al.* (1996); these experiments are discussed later in relation to mismatch repair processes). A potential problem with the proposed scheme is highlighted by the experiments of Ray *et al.* and Nickoloff *et al.*; repair of the double-strand break may result in crossing-over between the donor and acceptor genes, with loss of the intervening DNA (through the formation of an unstable plasmid, or by means of a nonconservative reciprocal exchange mechanism [Fishman-Lobell *et al.*, 1992; Santos-Rosa and Aguilera, 1994]). The frequency of such events, which can be as much as 50% for distant loci, are substrate-dependent: depending on the length/sequence of the duplicated and intervening region, and the chromosomal location of the duplication. A special case (as discussed previously) is presented by *MAT* and the *HM* loci, between which crossing-over rarely occurs. In order to guard against such deleterious reciprocal exchange events here, the *HIS4* gene, located between the synthetic donor/acceptor gene loci (Figure 1.19), will be used as a selectable marker. Its loss would mean that, under selective pressure, the host would die.

The donor and acceptor genes

The genes chosen on which to perform DNA shuffling by gene conversion, *OR17-40* and *OR17-228*, are two members of the human odorant receptor (OR) family. These putative odorant receptor genes were cloned from chromosome 17 by Ben-Arie *et al.* (1993), as part of an olfactory receptor gene cluster, comprising 16 genes in one 0.35 M.b. contiguous sequence.

Odorant receptors are presumptive G-protein coupled receptors that are expressed in olfactory sensory neurones in the mammalian nose (Buck, 2000). Several of their features make them attractive targets for DNA shuffling here: **1**. Like antibodies, they are involved in molecular recognition (albeit of small molecules, <310 daltons), and are capable of detecting millions of different molecular species,

often at extremely low concentrations (Mombaerts, 1999). Because of this extraordinary molecular recognition capability, odorant receptors may have some use in biosensor devices, e.g. for the detection of narcotics or explosives. **2**. Odorant receptor genes have intronless open reading frames that are approximately 1 k.b.p. in length. Further, they are homologous, with sequence identity ranging from 45% to >80% (Dryer and Berghard, 1999), making them ideal substrates for DNA shuffling. Also, the gene family is extremely large, estimated between 500 - 1000 OR genes in rat and mouse, and 500 - 750 in humans; surpassing the complexity of the immunoglobulin and T-cell antigen receptor gene families. **3**. The genes are entirely heterologous to the yeast genome, where no counterparts have been identified.

Because of the biosensor potential of the odorant receptors, the repertoire may be required for frequent selection against different targets. This might therefore be a good instance of where the gene conversion strategy for creating an appropriate repertoire might come into its own.

The genes of interest here (*OR17-40* and *OR17-228*) are members of the same subfamily (>80% identity), possessing 85.6% homology over their 951 b.p. open reading frames. In the experimental system, the *OR17-40* ORF will be integrated at the *LEU2* locus, and the *OR17-228* ORF at the *HML* locus, as shown in Figure 1.19. The methods used to integrate the genes are described in their relevant sections (Chapter 3). Note that the *OR17-228* ORF at the *HML* locus is flanked by sequences (~400 b.p.) homologous to the sequences found at either end of the *OR17-40* ORF at the *LEU2* locus (discussed below). In this first experiment it is not intended that the genes be expressed - no promoter elements have been incorporated - of concern here only is the efficacy of the gene conversion reaction.

The rare cutter endonuclease, I-SceI

The I-*Sce*I endonuclease (originally the *omega* transposease) has been termed a meganuclease, as it is one of only a few endonucleases that recognise a very long restriction sequence (Thierry and Dujon, 1992). The sequence of the unsymmetrical 18 b.p. recognition site was first determined by Colleaux *et al.* (1988a) and is shown in Figure 1.19. Cleavage of the site results in staggered 4 b.p. cuts, with 3'-OH overhangs (Colleaux *et al.*, 1988b).

The endonuclease is encoded by an optional *Saccharomyces cerevisiae* mitochondrial intron, Sc LSU.1, of the 21S rRNA gene (Macreadie *et al.*, 1985;

Jacquier and Dujon, 1985). The endonuclease is responsible for mobility and propagation of the intron (which behaves as a transposable genetic element), initiating a double-strand break at the intron insertion site. By means of site-directed mutagenesis, Colleaux *et al.* (1986), created a universal code equivalent (necessary due to the unusual codon usage in mitochondria) of the 235 codon long I-*Sce*I ORF, allowing the protein to be expressed from a suitable plasmid. Thierry *et al.* (1991) demonstrated (by digestion of total DNA) that no I-*Sce*I sites occur naturally in the yeast genome. Further, the ability of the endonuclease to enter the yeast genome (presumed to be a consequence of its small size) and to generate a double-strand break at a nuclear site was shown by Plessis *et al.* (1992). It was also shown, and of importance here, that the repair events induced by I-*Sce*I are identical to those initiated by the *HO* endonuclease.

The I-*Sce*I rare cutter endonuclease has been used to investigate various aspects of recombination and DNA repair in yeast (Fairhead and Dujon, 1993; Galli and Schiestl, 1998; Ricchetti *et al.*, 1999). I-*Sce*I induced double-strand breaks also increase recombination frequencies in plant (tobacco plant), mammalian (mouse/human) and insect (*Drosophila*) cells (Puchta *et al.*, 1993; Rouet *et al.*, 1994; Brenneman *et al.*, 1996; Rong and Golic, 2000).

Mode of the DNA shuffling event

As mentioned above, gene conversion at the acceptor locus will be initiated by a double-strand break at the I-*Sce*I recognition site, by the I-*Sce*I endonuclease. These events are designed to mimic events at the *MAT* locus, whereby the *HO* endonuclease initiates gene conversion, resulting in switching of the mating-type genes (Figure 1.9). Other aspects of the experimental system are also designed to copy the mechanistic principles of the mating-type switch. Thus, the donor gene, *OR17-228*, is flanked by regions of homology (~400 b.p.) to the acceptor gene, *OR17-40*, at the *LEU2* locus. In the same way that the mating-type genes - found in the distinct **Ya** and **Yα** regions - are flanked by the homologous W, X and Z regions (comprising ~1.5 k.b.p. of homology for the W and X regions, and ~0.3 k.b.p. for the Z region). It is hoped however, that the gene conversion events at the synthetic donor/acceptor locus will differ from those at the *MAT* locus in one crucial aspect: whilst mating-type switching results in the **Ya** sequence being replaced by the entire **Yα** sequence (or vice versa), gene conversion at the synthetic acceptor locus will occur in a stochastic manner, the

difference resulting from the homologous nature of the donor and acceptor genes, i.e. unlike the situation with Y**a** and Y**α**, exonuclease degradation of the cut strands will not have to proceed as far as the flanking regions before finding homology from which to initiate repair, and may in fact be initiated at any tract of homology between the two genes.

Gene conversion of the acceptor locus proceeds from the I-*Sce*I site, as shown in Figure 1.20. As described above, the extent of the conversion reaction, in any given induced cell, is unknown. Within an induction culture, potentially 10^8 cells, every possible conversion event (those allowed by the mechanics of the repair processes) might be expected to occur. After just one round of shuffling (one gene conversion event) it is conceivable that a receptor may have evolved to recognises a new target ligand. For instance, it was found that a mouse odorant receptor (I7), unlike the rat receptor, as a result of a single valine-to-isoleucine substitution, prefers heptanal instead of octanal (Krautwurst *et al.*, 1998). The result of several rounds of gene conversion (when multiple donor genes are present) is the creation of a mosaic acceptor gene, similar to that described for the antigenic surface proteins of *B. hermsii* and *T. brucei*, illustrated in Figure 1.2.

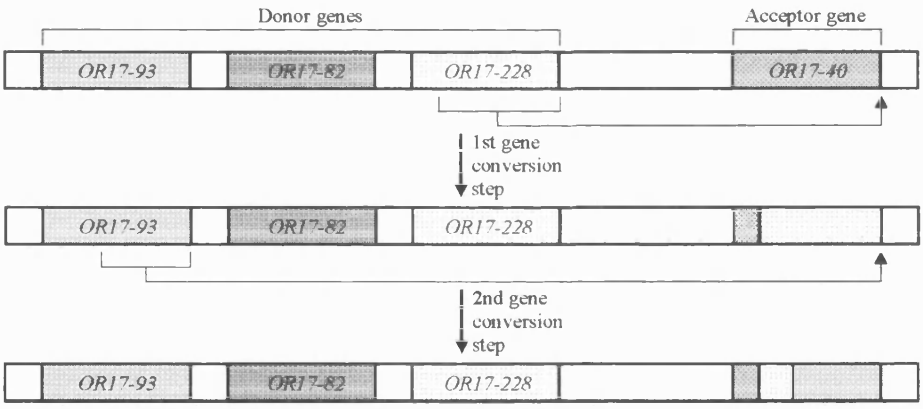


Figure 1.20 A simplified representation of the synthetic donor/acceptor locus. In this instance three donor genes are present, each of which can be used as a template for repair of the acceptor gene. Following the first repair event, ~80% of the acceptor gene has been converted to the *OR17-228* sequence, it is worth remembering that this represents the event in only one cell in an induced culture; other cells may have used 'more' or 'less' of the *OR17-228* gene, or have utilised either of the other two donor genes in the same way, from which to repair the acceptor gene. A colony selected showing improved ligand binding from the first round is then used as the seed strain for the second round of DNA shuffling; a 'mosaic' receptor gene is created when repair is initiated from a different donor gene, e.g. *OR17-93*, as shown.

Mismatch Repair

One potential inherent problem with the experimental system - crossing-over - has already been discussed. However, a second potential barrier to efficient repair of the acceptor gene (following induction) also exists. It has been shown (in yeast) that homeologous sequences (similar but not identical; Carpenter, 1984), containing 27% sequence divergence, recombine at about 5 - 10% of the frequency observed for homologous sequences (Bailis and Rothstein, 1990). Further, it has been shown that the presence of only a few mismatches (0.1% sequence divergence) can discourage recombination by as much as 50% (Borts *et al.*, 1990). This effect was first demonstrated in *Escherichia coli* (*E. coli*), where recombination between sequences that diverge by 10% was found to be 1000-fold lower than between homologous sequences (Rayssiguier *et al.*, 1989; Shen and Huang, 1989). However, in mutants defective for mismatch repair (*mutS*, *mutL*, and *mutH* mutants), the frequency of recombination events increase to nearly the frequency observed between homologous sequences. The regulation of recombination between diverged sequences is hypothesised to play several important roles, e.g., in maintaining integrity and diversity within gene families, by acting as a balancing force between mutation and conversion; in chromosome stability, preventing ectopic recombination between repeated sequences; and in speciation, reducing gene flow between diverging chromosomes (Alani *et al.*, 1994; Hunter *et al.*, 1996). Homologues of the bacterial genes *MutS* and *MutL* have been identified in organisms ranging from yeast to man. *PMS1* (Kramer *et al.*, 1989) and *MSH2* (Reenan and Kolodner, 1992) are yeast homologues of the bacterial mismatch repair genes *MutL* and *MutS*, respectively. Identified originally by the dramatic increase in frequency of post-meiotic segregation (PMS) of genetic markers, in *pms1* and *msh2* mutants (Williamson *et al.*, 1985; Reenan and Kolodner, 1992). This being indicative of unrepaired heteroduplex, and suggesting a role for these genes in the process of gene conversion.

The effects of null mutants of *PMS1* and *MSH2* upon the frequency of mitotic recombination between directly repeated homeologous sequences (Bailis and Rothstein, 1990; Selva *et al.*, 1995; Datta *et al.*, 1996), and meiotic recombination between homeologous chromosomes (Resnick *et al.*, 1992; Hunter *et al.*, 1996), has been investigated. The investigations of Datta *et al.* (1996) and Hunter *et al.* (1996) found a 10-fold enrichment effect for *pms1* recombination, in sequences that diverged by 8 - 20%. Although other workers, using even more diverged sequences, e.g. Bailis

and Rothstein (1990), Resnick *et al.* (1992), did not find a clear effect for *pms1*. However, in these studies *Msh2*, and a second *MutS* homologue, *Msh3*, were shown to increase the frequency of recombination. It is suggested by Hunter *et al.* (1996) that the method of processing recombination intermediates depends upon the degree of divergence between the participating molecules. Where, at relatively high levels of divergence (10 - 30%) *MSH2* appears to have a greater role than *PMS1* in preventing homeologous recombination.

Because of the very clear possibility that recombination will be reduced between the odorant receptor substrates (~15% diverged), an experimental strain will be constructed containing null mutations (for the reasons described above) of the mismatch repair genes. In the first instance (for the initial investigation) a *pms1* knock out strain will be created, for comparison against a wild-type experimental strain. Further work will involve the creation of *MSH2, MSH3*, and potentially multiple knock-out mutants (as evidence suggests that the effects of the genes may be additive).

A mismatch defective strain, if proven to be more effective than the wild-type strain in recombination, may itself, due to its nature, pose a problem. In this strain, mis-matched bases will not be corrected during the intended gene conversion reaction. This could lead, via multifarious repair processes (reviewed by Pâques and Haber, 1999), to co-conversion of the donor gene repertoire, illustrated in Figure 1.21.

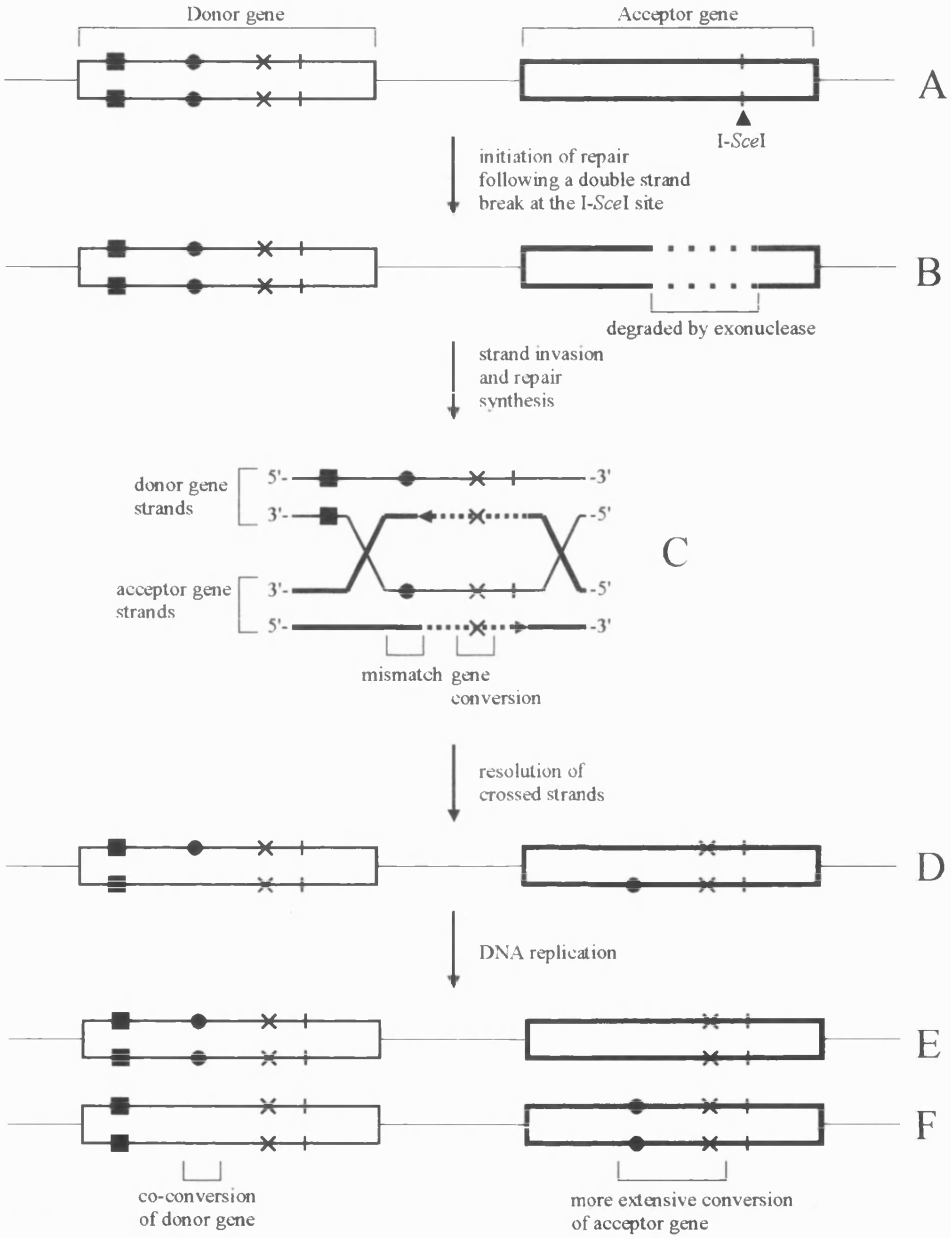


Figure 1.21 Mechanism leading to co-conversion of the donor gene in a mismatch repair deficient strain. **A**. The donor and acceptor gene are represented by medium and thick lines, respectively. Intervening DNA is shown as a thin line. Vertical bar: I-*Sce*I recognition site, and boundary between homologous sequence (to the right), and homeologous sequence (to the left). Mutations: square, circle and cross. **B**. A double-strand break at the I-*Sce*I site results in strand degradation via exonuclease activity (for simplicity here, both strands are shown as being degraded uniformly; discussed in Section 1.1.1). **C**. Repair synthesis of acceptor gene, leading to incorporation of the 'cross' mutation via gene conversion. A base pair mis-match (circle) within the heteroduplex DNA region is not corrected. Resolution, via symmetrical cleavage (crossed-strands in this example), of the Holliday junctions, leads to structure **D**, resulting in mis-matches at both donor and acceptor loci. DNA replication leads to the structures shown in **E** and **F**. In **F**, DNA repair has resulted in co-conversion of the donor gene, and a loss of sequence diversity.

Monitoring gene conversion

Gene conversion of the acceptor locus will be demonstrated by loss/gain of a unique restriction site (or sites): the acceptor gene will be amplified via PCR from selected colonies, digested by restriction enzyme, and the bands analysed by agarose gel electrophoresis (described in detail in Chapter 4). Colonies showing evidence of the desired reaction will be further analysed by DNA sequencing.

Future work

The eventual aim of this project is to create an *in vivo* DNA shuffling system, upon which multiple rounds of selection can be performed. Here, the first step in the construction of such a system, the creation, in a suitable host organism, of a synthetic donor/acceptor locus, that can be induced for gene conversion, is described. Future work, following successful demonstration of gene conversion in either a wild-type, or a mis-match repair defective strain (*pms1*, or a multiple knockout strain; discussed above), would involve the creation of a strain containing a repertoire of donor, odorant receptor genes (such a system would have to be studied carefully for any bias in the repair of the acceptor gene). At this stage the project would coincide with work done by other members of the group, involving expression of odorant receptors and their coupling of to signalling pathways. In this way, the binding of a ligand to an expressed odorant receptor would result in activation of a reporter gene, e.g. the green fluorescent protein of *Aequorea victoria* (Niedenthal *et al.*, 1996); thus allowing the crucial step of selection to be performed.

Other work

The work pertaining to that introduced in this chapter is described in Chapters 3 and 4. A second subsidiary project, investigating the efficacy of a Cre-based double recombination system to target DNA site-specifically into the yeast genome, is presented in Chapter 5.

CHAPTER 2

GENERAL MATERIALS AND METHODS

The general experimental techniques used throughout this project are given in this chapter. Any materials and methods specific to a particular chapter are described within that chapter.

2.1 MATERIALS

2.1.1 Chemical reagents

Except where stated otherwise chemical reagents (analytical grade) were obtained from Sigma, Fisions Ltd. or BDH Ltd.

The water used throughout this project was Millipore Q-Plus filtered water.

2.1.2 Molecular biological enzymes and reagents

T4 DNA ligase, T4 polynucleotide kinase, β-agarase I and all restriction endonucleases were purchased from New England Biolabs Inc (NEB). DNA polymerases were obtained from Bioline Ltd. (*Taq* DNA polymerase, obtained as BIOTAQ™), and Roche (*Taq/Pwo* DNA polymerase mix, obtained as Expand™ High Fidelity PCR systems). Ultrapure 2'-deoxynucleoside 5'-triphosphates (dNTPs) were purchased from Amersham Pharmacia Biotech as was [γ-^{32}P]ATP, obtained as 5'-[γ-^{32}P]triphosphate, triethylammonium salt (250 μCi, 10 mCi/ml, ~3000 Ci/mmol). Ultrapure™ agarose, Ultrapure™ low melting point agarose and 1 k.b.p. Plus DNA ladder were supplied by Life Technologies. Lyticase (5,000 - 20,000 units/mg protein; catalogue number L2524) and 5-fluoroorotic acid (5-FOA) were obtained from Sigma as were salmon sperm DNA, mineral oil and pancreatic RNase A.

2.1.3 Culture media and antibiotics

Bacto-agar, Bacto-yeast extract, Bacto-peptone, Bacto-tryptone, and Bacto-yeast nitrogen base without amino acids were all purchased from Difco Laboratories. The antibiotics ampicillin and geneticin (G418), the sugars glucose and galactose (minimum 99% pure), the base uracil and all amino acids were purchased from Sigma.

2.1.4 Other materials

Nylon membrane was obtained as MCI magna charge nylon transfer membrane from Genetic Research Instrumentation. Autoradiography film was obtained as X-Omat AR from Sigma, and 3MM paper was obtained from Whatman Scientific Ltd.

2.2 MOLECULAR BIOLOGICAL METHODS

Unless stated otherwise centrifugation was done in an MSE Micro-Centaur microfuge.

2.2.1 Design of oligonucleotide primers

Primers for amplifying from *S. cerevisiae* genomic DNA were designed with reference to DNA sequences obtained from the *Saccharomyces* Genome Database at Stanford (http://genome-www.stanford.edu/Saccharomyces/). Primers used to amplify from genomic DNA were at least 20 nucleotides in length. Primers used to amplify from vector DNA were shorter, typically 18 nucleotides in length (due to the lower sequence diversity of the vector). General considerations when designing primers included: choosing a target site with good base diversity; ensuring that the (G + C) base content of the primer was at least 50% to ensure stability of the primer-template duplex; and checking primer pairs for regions of complementarity to each other in order to avoid 'primer dimer' formation, which can lead to a reduction of the amount of available primers within a reaction.

If a restriction site was required at the end of a PCR product, 6 - 10 extra nucleotides were added after the enzyme recognition site to increase the efficiency of cutting. The reference section in the latest NEB catalogue "Cleavage Close to the End

of DNA Fragments" was consulted to decide the number and sequence of bases to flank the recognition site.

The sequences and descriptions of individual primers are given in their relevant chapters.

2.2.2 Synthesis of oligonucleotide primers

Oligonucleotides were synthesised by Perkin-Elmer Applied Biosystems and were supplied in 20% (v/v) acetronitrile/water and stored at -20°C.

An aliquot of each oligonucleotide was used throughout as a working stock to prevent contamination and repeated freeze-thawing. Oligonucleotides were used as supplied (i.e. as 20% (v/v) acetonitrile/water solutions).

2.2.3 The Polymerase Chain Reaction (PCR)

All PCRs were carried out in an M. J. Research Inc PTC-100 programmable thermal controller. The precise cycling parameters for individual reactions were optimised according to the following guidelines (based on the Expand™ High Fidelity PCR Systems information sheet supplied by Roche):

- DNA denaturation was carried out at 94°C.
- The annealing temperature used depended on the melting temperature of the primers used. If a PCR product was not obtained, the annealing temperature was decreased. If any non-specific products were observed the annealing temperature was increased.
- For amplifications up to 3 k.b.p. an elongation temperature of 72°C was used. For amplification of sequences longer than 3 k.b.p. an elongation temperature of 68°C was used.
- Elongation times were based on the below table:

Elongation time	45 s	1 min	2 min	4 min	8 min
PCR fragment length (k.b.p.)	up to 0.75	1.5	3	6	10

- To reduce the possibility of replicative errors the number of cycles was kept to a minimum.

Typically, analytical PCRs - looking for the presence of a particular DNA element - were done using BIOTAQ™ *Taq* polymerase (error rate 2.85 × 10^{-8}errors/b.p./cycle). Where greater fidelity was required - such as in the amplification of fragments for sequencing purposes or for transformation - Expand™ DNA polymerase was used. Expand™ DNA polymerase has a lower error rate (8.5 × 10^{-6} errors/b.p./cycle) due to the 3' - 5' exonuclease proofreading activity of the *Pwo* polymerase.

To check for contamination of PCR reagents a negative control (no template DNA) was done for each reaction; any bands in the negative control reaction would indicate contamination. For problematic reactions, where no product was obtained, a positive control (template DNA producing a known product) was included to show viability of reagents.

For reactions done in the thermal cycler without the heated lid one drop of mineral oil was added to each tube to prevent evaporation.

All reactions were analysed by agarose gel electrophoresis (Section 2.2.4).

A) *Taq polymerase chain reactions*

The reagent amounts for *Taq* DNA polymerase amplifications were: Each primer (25 pmol) and 5× *Taq* buffer (10 µl) [80 mM $(NH_4)_2SO_4$, 335 mM Tris-HCl (pH 8.8), 0.05% (v/v) Tween 20, 100 µM dATP, 100 µM dCTP, 100 µM dGTP, 100 µM dTTP, 7.5 mM $MgCl_2$]. These reagents were mixed in a thin-walled PCR tube (0.2 ml) with 50 - 250 ng of template DNA. BIOLINE™ *Taq* polymerase (0.5 µl) [5 units/µl] was added and the mixture was made up to a final volume of 50 µl with water.

The generalised PCR programme used for a *Taq* polymerase amplifications is given below:

1. Template denaturation (94°C for 2 minutes).
2. Template denaturation (94°C for 30 seconds).
3. Primer annealing (usually at 40 - 65°C for 30 seconds).
4. Elongation (72°C for 45 seconds to 8 minutes).
5. Steps 2 to 4 are repeated for 25 to 30 cycles.
6. Prolonged elongation (up to 7 minutes).

B) *Expand™ polymerase chain reactions*

The reagent amounts for an Expand™ DNA polymerase amplification were: Each primer (25 pmol), 10× Expand™ buffer (5 µl) [20 mM Tris-HCl (pH 7.5), 100 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA, 0.5% (v/v) Tween 20, 0.5% (v/v) Nonidet P40, 15 mM $MgCl_2$], 200 µM dATP, 200 µM dCTP, 200 µM dGTP and 200 µM dTTP. These reagents were mixed in a thin-walled PCR tube (0.2 ml) with 50 - 250 ng of template DNA. Expand™ DNA polymerase (0.75 µl) [3.5 units/µl] was added and the mixture was made up to a final volume of 50 µl with water.

The generalised PCR programme used for Expand™ polymerase amplifications is given below:

1. Denaturation (94°C for 2 minutes).
2. Denaturation (94°C for 15 seconds).
3. Annealing (usually at 40 - 65°C for 30 seconds).
4. Elongation (72°C for 45 seconds to 8 minutes).
5. Steps 2 to 4 are repeated for 10 cycles.
6. Denaturation (94°C for 15 seconds).
7. Annealing (usually at 40 - 65°C for 30 seconds).
8. Elongation (72°C for 45 seconds to 8 minutes with cycle elongation of 20 seconds per cycle for each subsequent cycle).
9. Steps 5 to 7 are repeated for 15 to 20 cycles.
8. Prolonged elongation (up to 7 minutes).

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis permits separation of DNA fragments by size exclusion. When a voltage is applied across the agarose gel the negatively charged DNA fragments migrate towards the positive terminal. The smaller the DNA fragment the faster it can migrate through the gel matrix; DNA fragments of different size therefore migrate at different speeds and are separated accordingly. The percentage of agarose in the gel is dependent on the size of the DNA fragments to be separated (or analysed). In this project 1% (w/v) agarose gels were most routinely used. 1% agarose gels permit good separation of DNA fragments between 500 b.p. and 4 k.b.p.

For analysis of DNA fragments <500 b.p., 1.5% (w/v) agarose gels were used. For DNA fragments between 4 k.b.p. and 10 k.b.p., 0.8% (w/v) agarose gels were used. Agarose gels were made up as follows: Ultrapure™ agarose (0.4 g - 0.75 g) [depending on percentage gel required] was dissolved in TBE buffer (50 ml) [90 mM Tris-borate, 2 mM EDTA] by boiling in a microwave oven (750W, full power for 2 min). Ethidium bromide was added to a final concentration of 0.5 μg/ml. Ethidium bromide is included in the gel for visualisation of the DNA fragments under ultra violet (UV) light. Gels were prepared in Bio-Rad™ Mini-sub® Cell Gel Trays. After setting, the gels were submerged in TBE buffer.

Typically 2.5μl, 3μl or 5μl of PCR product was loaded onto a gel for analysis. Prior to loading on the gel, the samples were mixed with 6× gel loading buffer (1 μl) [0.25% Bromophenol blue, 0.25% Xylene cyanol, 40% (w/v) sucrose in water]. The purpose of the loading buffer was to add colour and increase the density of the sample, allowing easier handling and ensuring that the sample dropped evenly into the well of the gel. The dyes in the loading buffer move at predictable rates towards the anode in the electric field, allowing visual inspection of how far the gel has run. Bromophenol blue migrates at the same speed as linear dsDNA 300 b.p. in length, whilst Xylene cyanol migrates at the same speed as linear dsDNA 4 k.b.p. in length (Sambrook *et al.*, 1989). This relationship holds for the range 0.5% to 1.4% agarose. Gels were typically run for 45 minutes at 70 volts (~7.0 V/cm).

To allow the size and concentration of DNA fragments to be determined the samples were run alongside a DNA ladder - the 1 k.b.p. Plus Marker (Figure 2.1). Typically 3 μl of marker (made up to a concentration of 0.16 μg/μl) was run on a gel. Estimates of sample DNA concentration were made by comparison of sample band brightness with that of the 1,650 b.p. marker band, which contains 10% of the total mass added to the gel (corresponding to 48 ng of DNA for 3μl loaded).

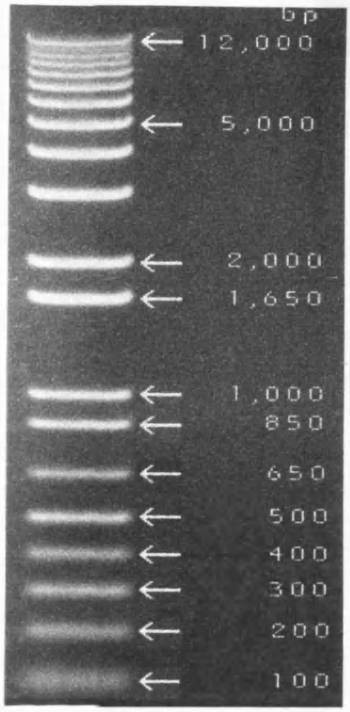


Figure 2.1 The 1 k.b.p. Plus Marker.

A) *Purification of DNA samples using LMP-agarose gels*

DNA fragments to be purified were resolved on low-melting point (LMP) agarose gels, from where the desired fragment could be excised and recovered.

LMP-agarose gels were prepared by dissolving Ultrapure™ LMP-agarose (0.5 g, for a 1% w/v gel) in TAE buffer (50 ml) [40 mM Tris-acetate, 1 mM EDTA] by heating in a microwave oven. Ethidium bromide was added to a final concentration of 0.5 μg/ml. Gels were prepared at 4°C in the cold room in Bio-Rad™ Mimi-sub® Cell Gel Trays. After forming gels were submerged in TAE buffer.

So that all of one DNA sample could be loaded into a single well of a gel samples were concentrated by alcohol precipitation (Section 2.2.5), typically to a final volume of 20 μl. Prior to loading, the samples were mixed with an appropriate volume of 6× LMP-gel loading buffer (0.25% Bromophenol blue, 40% (w/v) sucrose in water). Gels were run at 4°C for several hours at 40 volts (~4.0 V/cm). To ensure the correct size DNA fragment was excised from the gel, samples were run alongside 1 k.b.p. Plus Marker (5 μl) [0.16 μg/μl]. When possible, samples were visualised with a long wavelength (366 nm) hand-held UV lamp to avoid DNA damage. DNA

fragments were cut from the gel using single edged razor blades. A new razor blade was used for each DNA fragment.

Two techniques were used to recover the DNA of interest from an excised LMP-agarose gel slice: digestion with β-agarase I (next Section) and DNA extraction kits (Section 2.2.4c).

B) *Purification of DNA from LMP-agarose using β-agarase*

Gel slices cut from LMP-agarose gels were melted at 75°C, then 10× β-agarase buffer added to a final concentration of 10% total volume of the gel slice (the volume of the gel slice was determined approximately by weighing). The gel slice was allowed to cool to 40°C then β-agarase I was added to a final concentration of 2 units/200 µl of 1% LMP-agarose gel. The digest was incubated overnight at 40°C, then chilled on ice for 15 minutes in order to solidify any undigested agarose. Undigested agarose was removed from the digest by centrifugation in a microfuge (13 k.r.p.m for 15 minutes) at 4°C. The supernatant was transferred to a fresh polypropylene tube (1.5 ml) for alcohol precipitation (Section 2.2.5).

C) *Purification of DNA from LMP-agarose using DNA extraction kits*

Two different Promega DNA extraction kits were used to recover DNA from LMP-agarose. For DNA fragments larger than 3.5 k.b.p. the Wizard™ DNA Clean-Up System was used. For DNA fragments between 200 b.p. and 3.5 k.b.p. the Wizard™ PCR Preps DNA Purification Kit was used. The kits were used in accordance with the manufacturer's instructions. DNA was eluted from the extraction kit columns in water (35 µl), and stored at -20°C.

The DNA extraction kits were also used to clean PCR products directly in order to remove primers and PCR buffers so that the DNA could be used for subsequent reactions, such as restriction digests.

2.2.5 Alcohol precipitation of DNA

3M sodium acetate solution (pH 5.5) was added to the solution containing the DNA to be isolated to a final concentration of 300 mM. To this solution pre-chilled (-20°C)

isopropanol (2 volumes) was added. The solution was then incubated at -20°C for at least 1 hour. The DNA was pelleted by centrifugation in a microfuge (13 k.r.p.m. for 15 minutes) at 4°C. The supernatant was carefully removed so as not to disturb the DNA pellet. The pellet was left suspended in the remaining supernatant (~30 µl). The pellet was washed once in 70% ethanol (500 µl) then centrifuged again as above. The supernatant was removed as above to leave the DNA suspended in the remaining supernatant (~30 µl). The supernatant was then removed under vacuum by Speedivac (Savant Speed Vac Concentrator connected to Edwards Freeze Dryer and Vacuum Pump) for 20 minutes. Finally the dried pellet was resuspended in the required volume of water, and stored at -20°C.

2.2.6 *Escherichia coli* strain used and culture conditions

The *E. coli* strain used throughout this project was TOP10F' (F'{lacI^q *Tn*10(Tet^R)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139Δ(*ara-leu*)7697 *gal*U *gal*K *rsp*L *end*A1 *nup*G). The strain was obtained as part of the Eukaryotic TA Cloning Kit (Version 1.0) from Invitrogen. New strains made by transformation of TOP10F' with various cloning vectors are as described in the relevant chapters.

E. coli was cultured at 37°C in Luria broth (LB) [10 g/l Bacto-tryptone, 5 g/l Bacto-yeast extract, 10 g/l sodium chloride] liquid medium with shaking at 250 r.p.m., or on LB-agar plates (supplemented with 15 g/l Bacto-agar) overnight. For strains carrying plasmids conferring ampicillin resistance the medium was supplemented with ampicillin to a final concentration of 100 µg/ml.

2.2.7 *Saccharomyces cerevisiae* strains and culture conditions

Saccharomyces cerevisiae strains used are as described in their relevant chapters.

Strains were cultured at 30°C in yeast-extract, peptone, dextrose (YPD; Adams *et al.*, 1996) [10 g/l Bacto-yeast extract, 20 g/l Bacto-peptone, 20 g/l glucose, pH 5 to 6 with 3M NaOH] liquid medium with shaking at 250 r.p.m. overnight, or on YPD-agar plates (supplemented with 15 g/l Bacto-agar) for 2 to 3 days.

Strains with auxotrophic selectable markers were cultured at 30°C in Synthetic Complete (SC; Adams *et al.*, 1996) [(6.7 g/l Bacto-yeast nitrogen base without amino acids, 20 g/l glucose, 2 g/l 'drop-out' mix, pH 5 to 6 with 3M NaOH] liquid medium with shaking at 250 r.p.m. for 24 hours, or on SC-agar plates (supplemented with 20 g/l Bacto-agar) for 2 to 4 days. The SC medium was made up without the relevant amino acid(s)/base(s) in the drop out mix to select for the auxotrophic marker.

Drop out mix and other selective growth media used are described below:

A) *Drop out mix*

Drop out mix was made up as described by Adams *et al.* (1996):

L-Adenine	0.5 g	L-Leucine	10.0 g
L-Alanine	2.0 g	L-Lysine	2.0 g
L-Arginine	2.0 g	L-Methionine	2.0 g
L-Asparagine	2.0 g	L-Phenylalanine	2.0 g
L-Aspartic acid	2.0 g	L-Proline	2.0 g
L-Cysteine	2.0 g	L-Serine	2.0 g
L-Glutamine	2.0 g	L-Threonine	2.0 g
L-Glutamic acid	2.0 g	L-Tryptophan	2.0 g
Glycine	2.0 g	L-Tyrosine	2.0 g
L-Histidine	2.0 g	Uracil	2.0 g
L-Isoleucine	2.0 g	L-Valine	2.0 g

The constituents were thoroughly mixed by crushing with a pestle and mortar and then by turning end over end for at least 15 minutes. Drop out mix was stored in powder form in a light-proof container at room temperature until required.

B) *5-fluoro-orotic acid selective medium*

5-fluro-orotic acid (5-FOA) medium was prepared as described by Boeke *et al.* (1984): Yeast nitrogen base (7 g), 5-fluro-orotic acid (1 g), uracil (50 mg), glucose (20 g) were dissolved in water (500 ml) and filter sterilised (0.22 micron filter). The solution was mixed with 4 % autoclaved agar (500ml).

Where necessary, to complement specific auxotrophic requirements, amino acid(s) were added to the 5-FOA mix. The final concentration of the supplemented amino acid(s) was the same as in synthetic complete medium (Section 2.2.7).

C) *Sporulation medium*

Sporulation medium was prepared as described by Guthrie and Fink (1991):

Glucose	0.5 g	L-Histidine	20 mg
Yeast extract	2.5 g	L-Leucine	20 mg
Potassium acetate	15 g	L-Lysine	20 mg
Agar	20g	L-Methionine	20 mg
L-Adenine	40 mg	L-Tryptophan	20 mg
Uracil	40 mg	L-Phenylalanine	100 mg
L-Tyrosine	40 mg	L-Threonine	350 mg
		Water	1L

2.2.8 Preparation of glycerol stocks

Glycerol stocks were made of all *S. cerevisiae* and *E. coli* strains for long term storage. Overnight culture (800 µl) was added to 80% glycerol (200 µl), vortexed briefly, then stored at -70°C.

2.2.9 Transformation of *E. coli* cells with foreign DNA

Two methods - transformation by 'heat shock', and electrotransformation - were used at different stages in this project to transform competent *E. coli* TOP10F' cells. Both methods are described below. When carrying out a transformation a negative control was always done in which no DNA was added.

A) *Preparation of heat shock competent E. coli cells*

A single colony of TOP10F' was picked into LB (5 ml) [Section 2.2.6] and grown with shaking at 250 r.p.m. overnight until saturation. 500 µl of overnight culture was

inoculated into LB (50 ml) and grown at 37°C with shaking (250 r.p.m.) to an A_{450} of 0.4. The culture was then cooled on ice for 10 minutes before pelleting the cells by centrifugation (3 k.r.p.m. for 5 minutes) in a Sorvall GLC-4 general laboratory centrifuge, at 4°C. The pellet was resuspended in ice cold 0.1 M $CaCl_2$ (25 ml). The resuspended cells were then incubated for at least 1 hour on ice. The cells were pelleted again (as previously), and as much as possible of the supernatant discarded. Finally, the cells were resuspended in 2 ml of ice cold 0.1 M $CaCl_2$/18% (v/v) glycerol, snap frozen in aliquots (100 µl), and stored at -70°C. (Adapted from Sambrook *et al.*, 1989).

B) *Heat shock transformation of competent E. coli cells*

'Heat shock' competent *E. coli* TOP10F cells (100 µl) were gently thawed on ice. Their temperature was not allowed to go above 4°C. DNA [ligation mixture (5 µl) or plasmid (1 µl)] was added to the thawed cells and the cells incubated for 2 hours on ice. The cells were heat shocked at 42°C for 90 seconds, then returned to ice for 2 minutes. The cells were transferred to 2× TY (400 µl) [16 g/l Bacto-tryptone, 10 g/l Bacto-yeast extract, 5 g/l sodium chloride, pH 7.0 with 3M NaOH] containing no ampicillin, and incubated for 1 hour at 37°C with shaking. An aliquot (150 µl) of each transformation was then spread evenly onto 2× TY plates (2×TY medium supplemented with 15 g/l Bacto-agar) containing ampicillin (100 µg/ml). The plates were left to dry for 15 minutes, inverted, and incubated overnight at 37°C. (Adapted from Eukaryotic TA Cloning Kit instructions, Invitrogen).

C) *Preparation of electrocompetent E. coli cells*

L-broth (400 ml) [1% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl] was inoculated with 4 ml of fresh overnight culture. The cells were grown at 37°C with vigorous shaking to an A_{600} of 0.5 to 0.7. The cells were chilled on ice for 30 minutes then collected by centrifugation (4 k.r.p.m. for 15 minutes). As much of the supernatant as possible was removed and the cells resuspended in ice-cold sterile 10% (v/v) glycerol (400 ml), with care taken not to lyse the cells. The cells were centrifuged as previously and resuspended in ice-cold sterile 10% (v/v)

glycerol (200 ml). The cells were centrifuged again as previously and resuspended in ice-cold sterile 10% (v/v) glycerol (8 ml). The cells were centrifuged again (as previously) and resuspended in ice-cold sterile 10% (v/v) glycerol to a final volume of 800 µl. The cell concentration was in the range 1 - 3 $\times 10^{10}$ cells/ml. The cell suspension was frozen on dry ice in aliquots (40 µl) in polypropylene tubes (0.5 ml), and stored at -70°C. (Adapted from Pulse Controller Operating Instructions and Applications Guide, Bio-Rad.)

D) *High efficiency electro-transformation of E. coli cells*

An aliquot of electro-competent TOP10F' cells was gently thawed at room temperature and immediately placed on ice. An electroporation cuvette (0.1 cm) and the white chamber slide were placed on ice. The DNA to be transformed [ligation mixture (5 µl) or plasmid DNA (1 to 2 µl)] was mixed with the cell suspension (40 µl). The suspension was kept on ice until being transferred to the chilled electroporation cuvette. The Bio-Rad Gene Pulser apparatus was set to 25 µF and 1.80 kV. The Bio-Rad Pulse Controller was set to 200 Ω. The mixture of cells and DNA were transferred to the cold electroporation cuvette, and gently tapped to the bottom of the cuvette, ensuring that the sample was in contact with both aluminium sides of the cuvette. The cuvette was placed in the chilled safety chamber slide, and the slide pushed into the chamber until the cuvette was seated between the contacts at the base of the chamber. One pulse was applied at the above settings (a time constant between 4 and 5 ms was obtained). The cuvette was removed from the chamber and SOC medium (1 ml) [2% Bacto-tryptone, 0.5% Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose] immediately added, and the cells resuspended by gently pumping up and down with a pipette. The cell suspension was transferred to a 5 ml screw cap tube and incubated at 37°C with shaking (225 r.p.m. for 1 hour). Aliquots (150 µl) of the incubated cell suspension were spread evenly onto 2× TY plates (Section 2.2.9b) supplemented with ampicillin (final concentration of 100 µg/ml). The plates were allowed to dry for 15 minutes, inverted, and incubated overnight at 37°C. (Adapted from Pulse Controller Operating Instructions and Applications Guide, Bio-Rad.)

2.2.10 Transformation of *S. cerevisiae* cells with foreign DNA

Two methods - transformation using lithium acetate, and electrotransformation - were used at different stages in this project to transform *S. cerevisiae* cells. Both methods are described below. Cell densities were determined using a standard haemocytometer chamber (Section 2.2.11). When carrying out a transformation a negative control was always done in which no DNA was added.

A) *Preparation of lithium acetate competent S. cerevisiae cells*

A YPD (Section 2.2.7) preculture (10 ml) was inoculated from a single colony of the chosen *S. cerevisiae* strain and incubated overnight at 30°C with shaking. The cell density of the pre-culture was determined and diluted to 5×10^6 cells/ml in YPD (50 ml, final volume). The cells were grown for two additional cell divisions (~3 hours) and then recounted. The cell density was not allowed to exceed 1.7×10^7 cells/ml. The cells were collected by centrifugation (4 k.r.p.m. for 5 minutes) in a Sorvall GLC-4 general laboratory centrifuge. The cells were washed once in water (20 ml) and centrifuged again (4 k.r.p.m. for 5 minutes). The pellet was resuspended in water (1 ml) and transferred to a polypropylene tube (1.5 ml). The cell suspension was centrifuged in a microfuge (5 k.r.p.m. for 5 minutes), the supernatant removed and the pellet resuspended in freshly made 100 mM lithium acetate to a final cell density of 2×10^9 cells/ml. The cells were then incubated for 20 minutes at 30°C to make them fully competent.

B) *Transformation of lithium acetate competent S. cerevisiae cells*

Approximately 2.5 μg of the DNA to be transformed, and freshly boiled single-stranded salmon sperm DNA (10 mg/ml) [Section 2.2.19], was added to the lithium acetate competent cells (50 μl). The cells were vortexed briefly then incubated for 20 minutes at 30°C. A 40% (w/v) polyethylene glycol 3350/100 mM lithium acetate solution (300 μl) was added to the transformation mixture. The tubes were vortexed and incubated for 20 minutes at 30°C. The transformation mixture was heat shocked for 20 minutes at 42°C and the cells collected by centrifugation in a microfuge (5 k.r.p.m. for 1 minute). The supernatant was removed and the cells

resuspended in YPD (4 ml) and incubated with shaking for 2 hours at 30°C. The transformation culture was aliquotted into four polypropylene tubes and centrifuged in a microfuge (5 k.r.p.m. for 1 minute) and the supernatant removed. The cells were resuspended in water (500 µl) and aliquots (200 µl) inoculated onto appropriate plates for selection of transformants. The plates were incubated for 2 - 3 days at 30°C. (Adapted from Gietz and Woods, 1994; Wach *et al.*, 1996.)

C) *Preparation of electro-competent S. cerevisiae cells*

A YPD (Section 2.2.7) preculture (10 ml) was inoculated from a single colony of the chosen *S. cerevisiae* strain and incubated overnight at 30°C with shaking. The cell density of the pre-culture was determined and diluted to 2.5×10^7 cells/ml in YPD (50 ml, final volume). The cells were grown for two additional cell divisions (~3 hours) until a cell density of $\sim 1 \times 10^8$ cells/ml was reached. All subsequent cell washing steps were done using ice-cold solutions; centrifugation was done in a GLC-4 general laboratory centrifuge at 4°C. The temperature of the cells was not allowed to go above 4°C. The cells were collected by centrifugation (4 k.r.p.m. for 5 minutes) and resuspended in water (50 ml). Centrifugation was repeated (as before) and the cells resuspended in water (25 ml). Centrifugation was repeated (as before) and the cells resuspended in 1M sorbitol (1 ml). The cells were centrifuged for a final time (as before) and resuspended in 1M sorbitol (the resuspension volume was the same as the cell pellet volume, approximately 100 µl). The cells were kept on ice until they were used.

D) *High efficiency electro-transformation of S. cerevisiae cells*

An electroporation cuvette (0.2 cm) and the white chamber slide were placed on ice. In a polypropylene tube the electrocompetent *S. cerevisiae* cells (40 µl) were mixed gently with the DNA to be transformed [plasmid DNA (2 µl)[~0.1 µg], linear DNA (5 µl)[~0.1 µg]. The suspension was kept on ice for 5 minutes until being transferred to the chilled electroporation cuvette. The sample was tapped to the bottom of the cuvette, making sure that the sample was in contact with both aluminium sides of the cuvette. The Bio-Rad Gene Pulser apparatus was set at 25 µF and 1.50 kV. The

Bio-Rad Pulse Controller was set to 200 Ω. The cuvette was placed in the chilled safety chamber slide, and the slide pushed into the chamber until the cuvette was seated between the contacts at the base of the chamber. One pulse was applied at the above settings. A time constant between 4 and 5 ms was obtained for the pulse. The cuvette was removed from the chamber and ice-cold 1M sorbitol (1 ml) immediately added. The cells were resuspended by gently pumping up and down with a pipette. Aliquots (150 µl) of the cell suspension were spread evenly onto appropriate selective plates containing 1M sorbitol. The plates were allowed to dry for 15 minutes, inverted, and incubated for 2 - 3 days at 30°C. (Adapted from Pulse Controller Operating Instructions and Applications Guide, Bio-Rad.)

2.2.11 Counting yeast cells with a standard haemocytometer chamber

Yeast cell density was determined using a standard haemocytometer chamber (Weber Scientific International Ltd., supplied by Camlab). To determine cell density all the cells in a 1 mm^2 gridded area (volume 0.1 µl) were counted. Budding cells and cell clusters were counted as 1 cell (or 1 colony forming unit). Counting was done using the 40× microscope objective (with 10× ocular). To increase accuracy typically both chambers of the haemocytometer were counted. Cell density (cells/ml) was calculated using the formula below:

$$\text{Number of cells} \times 10^4 \times \text{dilution factor} = \text{cells/ml in culture}$$

The dilution factor was chosen according to the culture cell density. For exponential phase cultures (10^5 - 10^7 cells/µl) a dilution factor of 10 was used, for saturated cultures (10^8 - 10^9 cells/µl) a dilution factor of 100 was used.

2.2.12 PCR screening of transformants

In order to ascertain whether colonies growing on selective medium after transformation were *bona fide* positive colonies, i.e. contained the DNA element of interest correctly integrated (be it episomally or chromosomally), selected colonies where characterised by PCR screening. The methods used to rapidly screen *E. coli* for

episomal sequences, and *S. cerevisiae* for episomal and chromosomal sequences are described below. Glycerol stocks (Section 2.2.8) were made of any new strains.

A) *PCR screening of E. coli for vector sequences*

Typically ten colonies were screened at a time. Colonies were selected from the selective growth medium using a sterile cocktail stick and transferred to a PCR tube. The cocktail stick was then transferred to a numbered culture tube containing appropriate selective medium. The relevant *Taq* polymerase PCR mixture was then added to each PCR tube. PCR reagent amounts and cycling parameters were as given in Section 2.2.3a, except that the reaction volume was reduced to 20 µl and the initial denaturation step increased to 94°C for 5 minutes. The increased initial denaturation step was in order to disrupt the bacterial cells. PCR products were analysed by agarose gel electrophoresis, as described in Section 2.2.4.

B) *Rapid PCR screening of S. cerevisiae for episomal and chromosomal sequences*

In order to rapidly screen large numbers of yeast colonies 'whole cell' PCRs were performed (as described below). Where greater accuracy was required screening was done using genomic DNA, prepared as described in Section 2.2.13, and analysed by agarose gel electrophoresis (Section 2.2.4).

Colonies to be screened using the rapid screening method were numbered and then transferred, using a sterile cocktail stick, to a PCR tube containing water (~25 µl). The cells (a pin-head sized amount, ~5× 10^5 cells) were dispersed in the water by stirring. The cells were pelleted by centrifugation in a microfuge (13 k.r.p.m. for 1 minute) and the supernatant removed. The cells were then heated in a microwave oven (750W) at full power for 5 minutes. The cells were then suspended in the relevant *Taq* polymerase PCR mixture (as described in Section 2.2.3a, except that the reaction volume was reduced to 20 µl) by vortexing briefly or by pumping up and down with a pipette. After addition of the PCR mix the reaction was started as quickly as possible in order to prevent proteolysis, or the reaction was kept on ice until ready. PCR products were analysed by agarose gel electrophoresis (Section 2.2.4). [Adapted from Wach *et al.*, 1994.]

2.2.13 Preparation of yeast genomic DNA

A 10 ml culture, started from a single colony of the relevant *S. cerevisiae* strain, was grown to saturation overnight in YPD (Section 2.2.7)[or appropriate selective medium] at 30°C. The cells were collected by centrifugation (4 k.r.p.m. for 5 minutes) in a Sorvall GLC-4 general laboratory centrifuge. The supernatant was removed and the cells resuspended in water (0.5 ml). The cell suspension was transferred to a screw-capped polypropylene tube (1.5 ml) and centrifuged in a microfuge (13 k.r.p.m. for 30 seconds). The supernatant was removed and the cells resuspended in the residual liquid by vortexing. To the cell suspension, cell-lysis buffer (0.2 ml) [2% (w/v) Trition X-100, 1% (w/v) SDS, 100 mM NaCl, 10 mM Tris-HCl (pH 8.0)], phenol:chloroform:isoamyl alcohol (0.2 ml) [molar ratio 25:24:1; supplied by Life Technologies] and glass beads (0.3 g) [0.2 microns] were added. The mixture was vortexed vigorously for 4 minutes. Water (0.2 ml) was added, the mixture vortexed briefly then centrifuged in a microfuge (13 k.r.p.m. for 5 minutes). The top aqueous layer was transferred to a new polypropylene tube (1.5 ml) and absolute ethanol (1 ml) added to precipitate the DNA. The DNA solution was then vortexed briefly and centrifuged in a microfuge (13 k.r.p.m. for 2 minutes). The supernatant was discarded and the DNA pellet resuspended in water (0.4 ml). RNaseA (6 µl) [10 mg/ml; Section 2.2.18] was added and the solution incubated at 37°C for 30 minutes. Absolute ethanol (1 ml) and 4M ammonium acetate (10 µl) were added, the solution vortexed briefly and centrifuged in a microfuge (13 k.r.p.m. for 2 minutes). The DNA pellet was washed with 70% (v/v) ethanol and left to air-dry overnight. Finally the pellet was resuspended in water (50 µl), to give a final DNA concentration of ~200 - 400 ng/µl. (Adapted from Adams *et al.*, 1996.)

2.2.14 Restriction endonuclease digestion of DNA

All restriction enzymes were obtained from New England Biolabs (NEB) and used according to the manufacturer's instructions. Typically 1 - 1.5 µg of DNA was digested in a single reaction. Digestions were usually performed in a final reaction volume of 40 µl. The reaction mixture consisted of the required 10× NEB buffer (4 µl), and restriction enzyme (1µl) [3 - 20 units]. Where required, bovine serum albumin was added to the digest mixture, to a final concentration of 100 $\mu g ml^{-1}$. The

reaction mixtures were incubated at the required temperature for 16 hours and then heat inactivated at 65°C for 20 minutes.

2.2.15 Ligation of sticky ended DNA

Ligation mixtures were prepared using 6 Weiss units (400 NEBunits/µl) of T4 DNA ligase, 10× NEB ligase buffer (1 µl) [500 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$, 100 mM dithiothreitol, 10 mM ATP, 250 µg/ml bovine serum albumin], insert and vector DNA with compatible cohesive ends, in a final reaction volume of 10 µl.

The molar ratio of insert to vector used was approximately 10:1 (amount of DNA was estimated as described in Section 2.2.4). As a control to assess the degree of vector self-religation a 'cut vector only' reaction was prepared each time. Ligation mixtures were incubated at 16°C overnight in the thermal cycler and then heat-inactivated at 65°C for 10 minutes.

Ligation mixtures were transformed directly into competent *E. coli* TOP10F' cells (Section 2.2.9).

2.2.16 Automated DNA sequencing and sequence analysis

DNA sequencing was done on an Applied Biosystems PRISM™ 377 DNA Sequencer using the fluorescent dye terminator protocol, based on the chain-terminator sequencing method devised by Sanger *et al.* (1977). Samples were prepared in thin-walled PCR tubes (0.2 ml). Each sample contained DNA template (~150 ng), and primer (10 pmol). The sample was made up to a final volume of 6 µl with water. Analysis of resulting DNA sequence chromatographs was done using the Sequence Editor (SeqEd™) programme (Version 1.03, Perkin-Elmer).

DNA sequence manipulation

For routine manipulation of DNA sequences, such as restriction analysis, codon usage analysis and in the construction of recombinant fragments, the MacVector™ programme (Version 7.0, Molecular Simulations Ltd) was used.

2.2.17 Preparation of plasmid DNA

Plasmid DNA was prepared from an overnight culture of TOP10F' *E. coli* cells (grown in LB medium; Section 2.2.6) carrying the plasmid of interest using the S.N.A.P.™ Miniprep Kit (Invitrogen), according to the manufacturer's instructions. The DNA was eluted from the S.N.A.P.™ miniprep column with water (60 µl), and stored at -20°C. Plasmid concentration was determined using the formula below:

$$[\text{DNA}] = (A_{260})(0.05 \text{ mg/ml}) \times \text{dilution factor}$$

Plasmid DNA was typically found to be in the range 0.1 - 0.5 µg/µl.

2.2.18 Preparation of pancreatic RNaseA

Pancreatic RNaseA (100 mg) was dissolved in 10 mM Tris-HCl (10 ml, pH 7.5), 15 mM NaCl and heated at 100°C for 15 minutes. The mixture was allowed to cool to room temperature then dispensed into aliquots (0.5 ml), and stored at -20°C.

2.2.19 Preparation of salmon sperm DNA

Salmon sperm DNA (10 mg/ml) was prepared by dissolving salmon sperm DNA (0.2 g) in water (20 ml). The solution was stirred until the DNA had dissolved, the DNA was then fragmented by pushing the solution ~30 times through a small gauge needle. The fragmented DNA was aliquotted and stored at -20°C.

2.2.20 Southern blotting

DNA samples for Southern blotting (Southern, 1975) and oligonucleotide probes were prepared as described in relevant chapters. Transfer of DNA from the agarose gel to the membrane and the hybridisation protocol are described below:

A) *Agarose gel electrophoresis*

For Southern blotting a 0.8 % (w/v) agarose gel (Section 2.2.4) was prepared in a Bio-Rad™ Mimi-sub® Cell Gel Tray. The gel composition was: Ultrapure™ agarose

(1.6 g) dissolved in 0.5 × TBE buffer (200 ml). Ethidium bromide was added to a final concentration of 0.2 μg/ml. After setting, gels were submerged in 1 × TBE buffer containing ethidium bromide (0.1 μg/ml). To allow DNA fragment size to be determined 1 k.b.p. Plus Marker (5 μl)[Section 2.2.4] was run alongside the DNA samples.

B) *Transfer of DNA to nylon membrane*

Transfer of DNA from the gel to the nylon membrane (MCI Magna charge nylon transfer membrane) was as described by Sambrook *et al.* (1989), except at the final step the DNA was fixed to the membrane by UV-crosslinking. Crosslinking was done in a CL-1000 ultraviolet crosslinker using 1200 preset energy units (×100 $\mu J/cm^2$).

The crosslinked filter was stored at room temperature between two sheets of 3MM paper wrapped in aluminium foil until used.

C) *Hybridisation and pre-hybridisation*

The membrane to be probed was laid face upwards on gauze and rolled carefully so that the side to be probed was innermost. The rolled membrane was inserted into a glass screw-capped hybridisation tube.

Salmon sperm DNA (10 mg/ml; Section 2.2.19) was denatured by boiling for 5 minutes and immediately added to pre-hybridisation buffer (0.5 M Na_2HPO_4, 1 mM EDTA, 7 % SDS) to a final concentration of 100 μg/ml. The pre-hybridisation buffer (30 ml) was then added to the membrane. Pre-hybridisation was carried out in a Hybaid hybridisation oven at 42 °C for 2 hours, with the sample rotating slowly. The pre-hybridisation buffer was discarded and pre-heated hybridisation buffer (30 ml) added (the pre-hybridisation and hybridisation buffer were the same). Immediately prior to use, salmon sperm DNA (as for pre-hybridisation), and the radiolabelled oligonucleotide probe (30 μl), were added to the hybridisation buffer. Hybridisation was carried out overnight at 42°C, with the sample rotating slowly. The hybridisation solution was discarded and the membrane washed three times (as follows) using pre-heated solutions. Each wash was performed using as large a volume as possible. The first wash (5% SDS, 40 mM Na_2HPO_4, 1 mM EDTA) was at 42°C for 30 minutes, with the container slowly rotating. The second wash (1% SDS, 40 mM

Na_2HPO_4, 1 mM EDTA) was at 50°C for 2 hours, with the container slowly rotating. For the final wash (40 mM Na_2HPO_4) the membrane was transferred to an open container, where it was gently agitated at room temperature for 5 minutes.

Excess solution was allowed to run off the membrane, which was then placed face up on paper towels and allowed to dry. The membrane was monitored at this stage with a Geiger counter. The membrane was then carefully wrapped in Saran Wrap and exposed to photographic film (X-Omat AR) at -70°C. The exposed film was developed using an Amersham hyperprocessor.

Hybridisation and wash temperatures were chosen according to Tomlinson *et al.* (1992). The hybridisation protocol was adapted from Church and Gilbert (1984).

CHAPTER 3

DIRECTED MOLECULAR EVOLUTION BY GENE CONVERSION: CONSTRUCTION OF EXPERIMENTAL STRAINS

3.1 INTRODUCTION

The aims and the overall experimental strategy for the Molecular Evolution by Gene Conversion strategy were described in Chapter 1. This chapter describes the construction of the experimental strains (as outlined in Section 1.3) and various control strains (described below) that are required. The experiments performed with the strains, once constructed, are presented in Chapter 4.

Experimental strains

Two experimental strains are to be constructed, based on the model shown in Figure 1.19. The second strain differs from the first in that a gene involved in mismatch repair, *PMS1*, has been disrupted (for reasons described in Section 1.3). The main steps to be described - with reference to Figure 1.19 - are: **1**. integration of the acceptor gene, *OR17-40* (containing the I-*Sce*I recognition site) at the *LEU2* locus. **2**. integration of the donor gene, *OR17-228* (containing the I-*Sce*I recognition site, and flanked by regions homologous to the donor locus) at the *HML* locus. **3**. Construction of a vector capable of expressing the rare-cutter endonuclease, I-*Sce*I, under the control of the yeast *GAL1* promoter (described in Appendix A), and transformation of the vector into the experimental strains. Two control strains (one for each experimental strain) will also be constructed, where the open reading-frame of the I-*Sce*I endonuclease is absent from the galactose-inducible vector (described as 'vector-only' controls, where no endonuclease production can occur under any

conditions). Other steps to be described include, the conversion of the experimental strain to a *HIS*$^+$, *GAL*$^+$ genotype (for reasons described in appropriate sections), by transformation and mating experiments, respectively.

The four experimental strains (i.e., the two experimental strains and their respective 'vector-only' controls) are constructed in a step-wise manner from the parent strain, JRY50 (the reason for choosing this particular starting strain is described later; Section 3.3.3). A simplified overview of the main steps performed in the construction of the experimental strains is shown in Figure 3.1. The steps shown are described in detail in their relevant sections.

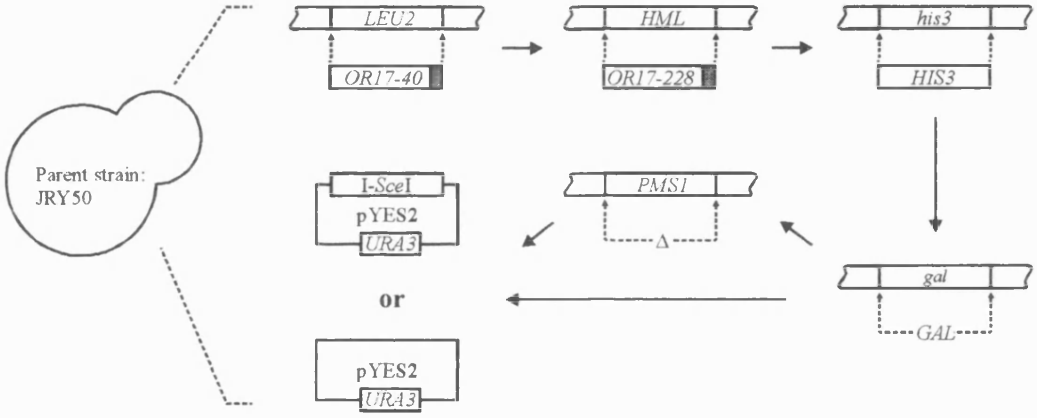


Figure 3.1 The main steps in the construction of the experimental strains from the parent strain, JRY50. The steps are shown in the order in which they were performed, and are described in the same order in following sections of this chapter. The means by which each step is carried out is described in detail in the appropriate section. The four experimental strains constructed are: **BGY11**, containing the complete experimental system; **BGY12**, the *pms1* (mismatch repair deficient) derivative of BGY11; **BGY13**, the 'vector-only' control for BGY11 (containing pYES2 only, instead of the I-*Sce*I expression vector pYES2[I-*Sce*I]); and **BGY14**, the 'vector-only' control for BGY12. Detailed descriptions of the genotypes of these strains (and intermediate strains) are given in Table 3.2.1. Filled rectangle: I-*Sce*I recognition site. [A more detailed summary of the steps presented above, and for the construction of the control strains described below, is given in Appendix B.]

Control strains

The construction of two different control strains is described. Each control strain is designed to look at a different aspect or component of the reaction being performed in the experimental strains (above). As for the experimental strains, each control strain

was also constructed with a 'vector-only' version (as above), giving a total of four strains.

The first control strain is based on the parent strain FY833 and is designed to show that I-*Sce*I expression is occurring as intended. The FY833 control strain is designated the 'vector control', as expression of the endonuclease, from pYES2[I-*Sce*I], is intended to cleave an I-*Sce*I recognition site on a second vector, pRS313[I-*Sce*I], rendering the strain *his3*$^-$, and unable to grow on medium lacking histidine. In this simple experiment, endonuclease production is to be demonstrated by reduced cell viability (on medium lacking histidine) following induction in galactose-containing medium (described further in relevant sections and in Chapter 4).

The second control strain, designated the 'gene conversion' control strain, is intended to act as a 'donor-deficient' control for the experimental strains. This strain - based on a strain constructed during the course of the work presented in Chapter 5 - is very similar to the experimental strain, BGY11, containing the *OR17-40* ORF (followed by the I-*Sce*I recognition site) integrated at the *LEU2* locus, but does not contain the donor gene (*OR17-228*) at the *HML* locus (which is wild-type). It is expected that endonuclease production in this strain will result in reduced cell viability due to a double strand break that cannot be repaired (described further in relevant sections and in Chapter 4).

A simplified overview of the steps performed in the construction of both control strains ('vector control' and 'gene conversion' control) is shown in Figure 3.2. The steps shown are described in detail in their relevant sections.

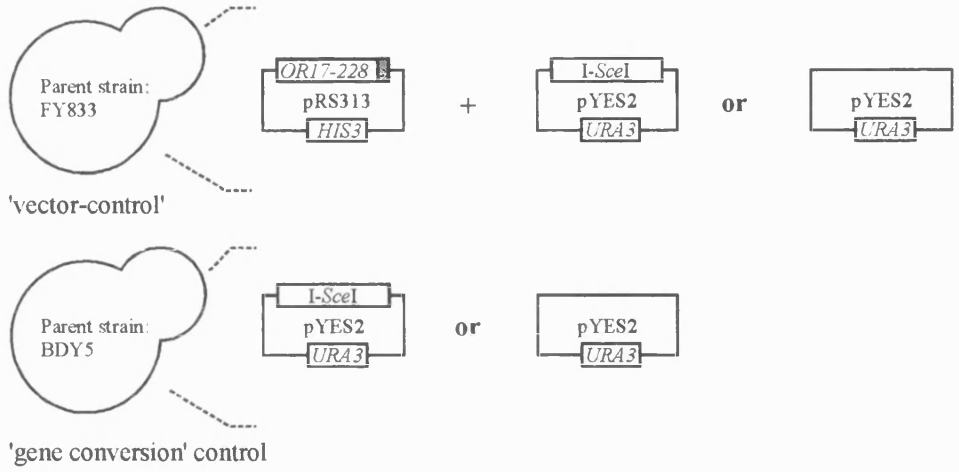


Figure 3.2 Construction of the 'vector control' and 'gene conversion' control strains (steps are described in detail in appropriate sections). The parental strain for both control strains is given on the left hand side. The two 'vector control' strains are generated by transforming FY833 with pRS313[I-*Sce*I] and pYES2[I-*Sce*I], to give **BGY17**, and with pRS313[I-*Sce*I] and pYES2 only, to give **BGY18**. The 'gene conversion' control strain - **BGY15** - is generated by transforming strain BDY5 (constructed during the course of the work presented in Chapter 5) with pYES2[I-*Sce*I], the 'vector-only' version of this strain - **BGY16** - is generated by transforming BDY5 with pYES2 only. Detailed descriptions of the genotypes of these strains (and intermediate strains) are given in Table 3.2.1. Filled rectangle: I-*Sce*I recognition site.

The steps shown in Figures 3.1 and 3.2 represent the main steps to be described in the construction of the experimental and control strains. At the beginning of each main step in the following Methods section an overview is presented, summarising what is to be described in that section.

3.2 MATERIALS

3.2.1 *Saccharomyces cerevisiae* strains

Supplied *Saccharomyces cerevisiae* strains, and those constructed during the course of the work in this chapter, are described herein.

Strain	Genotype	Reference
JRY50	*MAT***a** *his3-532 trp1-289 ura3-52 sir4-9*	Schnell and Rine, 1986.
FY833	*MAT***a** *his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63 GAL2*$^+$	Winston *et al.*, 1995.
FY3	*MAT***a** *ura3-52 GAL2*$^+$	Winston *et al.*, 1995.
INVSC2	*MAT*α *his3Δ200 ura3-52 GAL*$^+$	Supplied by Invitrogen®.
MR9	*MAT***a** *sst1Δ sst2Δ his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63 GAL2*$^+$	Redfern, 2000.
RC757	*MAT*α *sst2-1 rme his6 met1 can1 cyh2*	Supplied by A. Wheals.
BGY1	*MAT***a** *his3-532 trp1-289 sir4-9 leu2Δ::URA3*	This work.
BGY2	*MAT***a** *his3-532 trp1-289 ura3-52 sir4-9 leu2Δ::OR17-40*	This work.
BGY3	*MAT***a** *his3-532 trp1-289 sir4-9 leu2Δ::OR17-40 hmlΔ::URA3*	This work.
BGY4	*MAT***a** *his3-532 trp1-289 ura3-52 sir4-9 hmlΔ::URA3*	This work.
BGY5	*MAT***a** *his3-532 trp1-289 leu2Δ::OR17-40 sir4-9 hml::*pRS306[*OR17-228*]	This work.
BGY6	*MAT***a** *his3-532 trp1-289 ura3-52 sir4-9 leu2Δ::OR17-40 hmlΔ::OR17-228*	This work.
BGY7	*MAT***a** *HIS3 trp1-289 sir4-9 ura3-52 leu2Δ::OR17-40 hmlΔ::OR17-228*	This work.

continued next page.

Strain	Genotype	Reference
BGY8	*MAT***a** *HIS3 trp1-289 sir4-9 ura3-52 leu2Δ::OR17-40 hmlΔ::OR17-228 GAL*$^+$	This work.
BGY9	*MAT***a** *HIS3 trp1-289 sir4-9 ura3-52 leu2Δ::OR17-40 hmlΔ::OR17-228 GAL*$^+$ *pms1::*pRS306[*PMS1*Δ]	This work.
BGY10	*MAT***a** *HIS3 trp1-289 sir4-9 ura3-52 leu2Δ::OR17-40 hmlΔ::OR17-228 GAL*$^+$ *pms1*Δ	This work.
BGY11	BGY8::pYES2[I-*Sce*I] *URA3*	This work.
BGY12	BGY10::pYES2[I-*Sce*I] *URA3*	This work.
BGY13	BGY8::pYES2 *URA3*	This work.
BGY14	BGY10::pYES2 *URA3*	This work.
BGY7Z	BGY7::pYX243 *LEU2*	This work.
BGY8Z	BGY8::pYX243 *LEU2*	This work.
FY10Z	*MAT*α *ura3-52 leu2Δ1 GAL2*$^+$ pYX243 *LEU2*	This work.
BGY15	BDY5::pYES2[I-*Sce*I] *URA3*	This work.
BGY16	BDY5::pYES2 *URA3*	This work.
BGY17	FY833::pYES2[I-*Sce*I] *URA3* pRS313[I-*Sce*I] *HIS3*	This work.
BGY18	FY833::pYES2 *URA3* pRS313[I-*Sce*I] *HIS3*	This work.

The yeast strain JRY50 was chosen as the starting point from which to construct the experimental strain. The reason for choosing this particular strain is described in Section 3.3.3.

The yeast FY-strains represent the principal strains used in Chapter 5, and are described in Section 5.2.1.

3.2.2 Plasmids

Supplied DNA vectors used in this chapter are described herein. Plasmids generated during the course of the work in this chapter are described in their relevant sections.

A) *pYES2[OR17-40]*

Plasmid pYES2[*OR17-40*] was constructed by Redfern (2000). It contains the open reading frame of the putative human odorant receptor, *OR17-40*, cloned between the *Hin*dIII and *Xba*I restriction sites in the multiple cloning region of plasmid pYES2 (described below). The cloned gene is orientated such that the initiation codon is after the *Hin*dIII site.

B) *pYES2*

Plasmid pYES2 as supplied by Invitrogen® is shown in Figure 3.3. pYES2 is a yeast expression vector designed for high level expression in *S. cerevisiae*. The vector can be maintained in both yeast and *E. coli*. It contains the yeast *GAL1* promoter for inducible expression of genes inserted into the multiple cloning site, and the yeast *CYC1* transcription termination signal sequence. The auxotrophic marker for selection in yeast is the *URA3* gene. Other features are the ampicillin gene and *E. coli* origin from pUC19 as well as the T7 RNA polymerase promoter and f1 origin, for maintenance, selection, transcription and single strand rescue in *E. coli*.

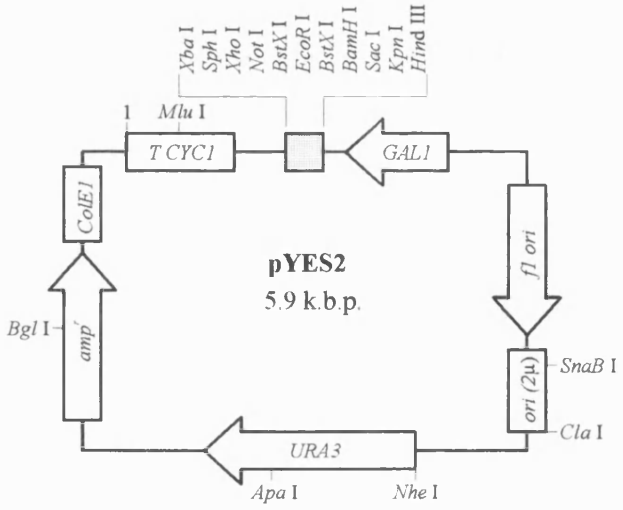


Figure 3.3 The yeast expression vector pYES2. Multiple cloning site (shaded box), unique restriction sites are shown (Figure is not to scale).

C) *pTG5756*

Plasmid pTG5756 (Figure 3.4), as constructed by Längle-Rouault and Jacobs (1995), was a gift from Transgene S.A. (Strasbourg). The plasmid was obtained by cloning the *Kluyveromyces lactis URA3* gene (accession number D00431, n.t. 151 - 1652) between the *Apa*I and *Not*I sites of pRS315 (Sikorski and Hieter, 1989).

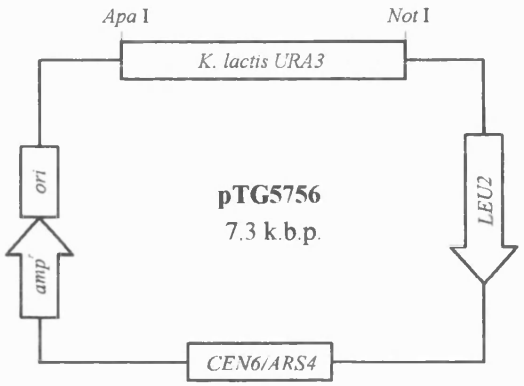


Figure 3.4 Plasmid pTG5756 containing the cloned *K. lactis URA3* gene (orientation unknown). Figure is not to scale.

D) *pRS306*

The yeast integrating plasmid pRS306 as constructed by Sikorski and Hieter (1989) is shown in Figure 3.5. The vector contains the ampicillin gene for selection in *E. coli* and the *URA3* auxotrophic marker for selection in yeast.

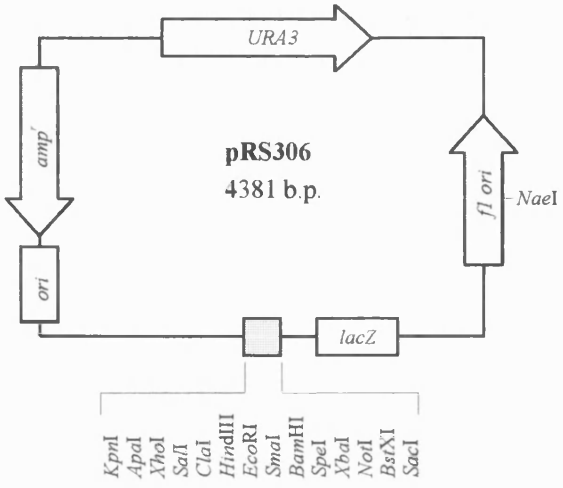


Figure 3.5 Plasmid pRS306. Multiple cloning site (shaded box), unique restriction sites are shown (Figure is not to scale).

E) *pPEX7*

The I-*Sce*I expression plasmid pPEX7 as constructed by Fairhead and Dujon (1993) is shown in Figure 3.6. The plasmid contains the universal code equivalent of the I-*Sce*I rare cutter endonuclease, under the control of the inducible *GAL10-CYC1* promoter.

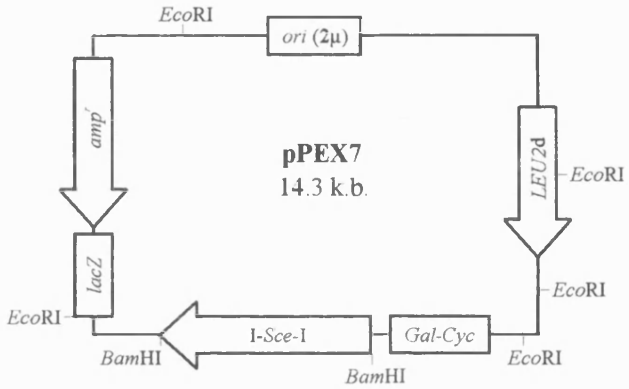


Figure 3.6 I-*Sce*I expression plasmid pPEX7. The plasmid is maintained at a high copy number (100 - 200 per cell) as the *LEU2*d allele poorly complements leucine auxotrophy (Guthrie and Fink, 1991). Figure is not to scale.

F) *pYES2[OR17-228]*

Plasmid pYES2[*OR17-228*] was constructed by Crickett (1997). It contains the open reading frame of the putative human odorant receptor, *OR17-228*, cloned between the *Hin*dIII and *Xba*I restriction sites in the multiple cloning region of plasmid pYES2 (Section 3.2.2b). The cloned gene is orientated such that the initiation codon is after the *Hin*dIII site.

G) *pYX243*

Plasmid pYX243 as supplied by R&D Systems is shown in Figure 3.7. The vector contains the *LacZ* gene under the control of the yeast *GAL1* promoter.

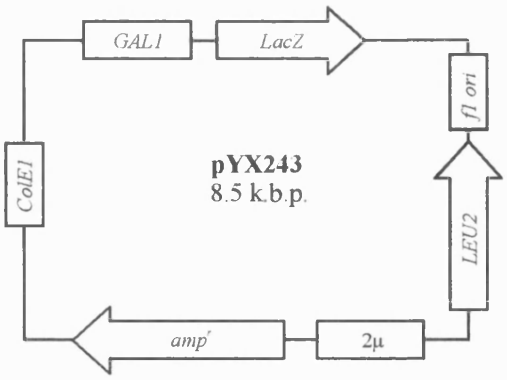


Figure 3.7 The β-galactosidase expression vector pYX243. The plasmid replicates autonomously due to the 2μ-sequence, and is maintained between 25 - 100 copies per cell (Figure is not to scale).

H) *pRS313*

The yeast centromeric plasmid pRS313 as constructed by Sikorski and Hieter (1989) is shown in Figure 3.8. The vector contains the ampicillin gene for selection in *E. coli* and the *HIS3* auxotrophic marker for selection in yeast.

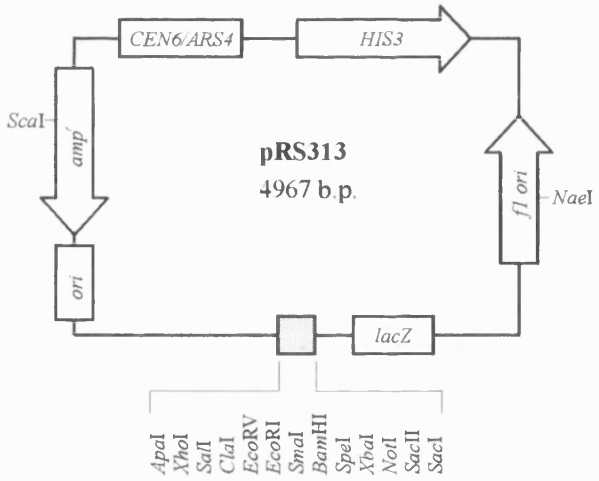
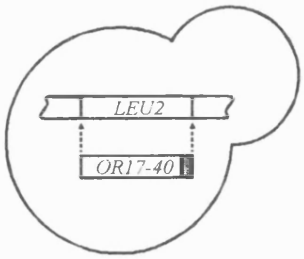


Figure 3.8 Plasmid pRS313. Multiple cloning site (shaded box), unique restriction sites are shown (Figure is not to scale).

3.3 METHODS: GENERATION OF EXPERIMENTAL STRAINS FOR THE ANALYSIS OF MOLECULAR EVOLUTION BY GENE CONVERSION

Overview: The next section describes the steps performed in order to integrate the *OR17-40* ORF (followed by the I-*Sce*I recognition site) at the *LEU2* locus in the parent strain, JRY50.



3.3.1 Construction of a cassette for the integration of *OR17-40* ORF at the acceptor locus *LEU2*

The yeast strain JRY50 (Section 3.2.1) was chosen as the strain in which to construct the experimental model system (Figure 1.20). The choice of this particular strain was dependent on the strategy to be used for the integration of the donor gene, the *OR17-228* ORF, at the *HML* locus; this is discussed in Section 3.3.3.

In this section the integration of the acceptor gene, the *OR17-40* ORF, at the *LEU2* locus is described. The strategy used to achieve this is based on the method described be Alani *et al.* (1987) and later developed by Längle-Rouault and Jacobs (1995), thus: a *URA3* gene is flanked by directly repeated sequences which contain homology to the desired target locus (*LEU2*), and the desired mutation (the *OR17-40* ORF). After introduction of the fragment into the yeast genome by transformation, stable integrants are isolated by selection for Ura^+ colonies. In order to minimize non-desired recombination events with genomic *ura3* sequences, the heterologous (~68% identical n.t.) *URA3* gene from the yeast *Kluyveromyces lactis* is used. Recombination between the flanking direct repeats, which occurs at a high frequency ($\sim 10^{-4}$) in vegetatively grown cultures, results in marker excision and the stable integration of one copy of the desired mutation. Derivatives that have

undergone the excision event are Ura$^-$, and can be selected directly with 5-fluoro-orotic acid (5-FOA)[Boeke *et al.*, 1984].

Two regions of homology to the target locus were incorporated into the flanking repeat sequences. A region homologous to the promoter region of the *LEU2* gene, and a region homologous to the terminator region of the *LEU2* gene. Each homologous region being approximately 300 b.p. in length. The *LEU2* sequences were separated by the desired mutation, the *OR17-40* ORF (with an I-*Sce*I restriction site after the termination codon). Successful cassette integration, followed by marker excision, should result in replacement of the *LEU2* ORF with the *OR17-40* ORF.

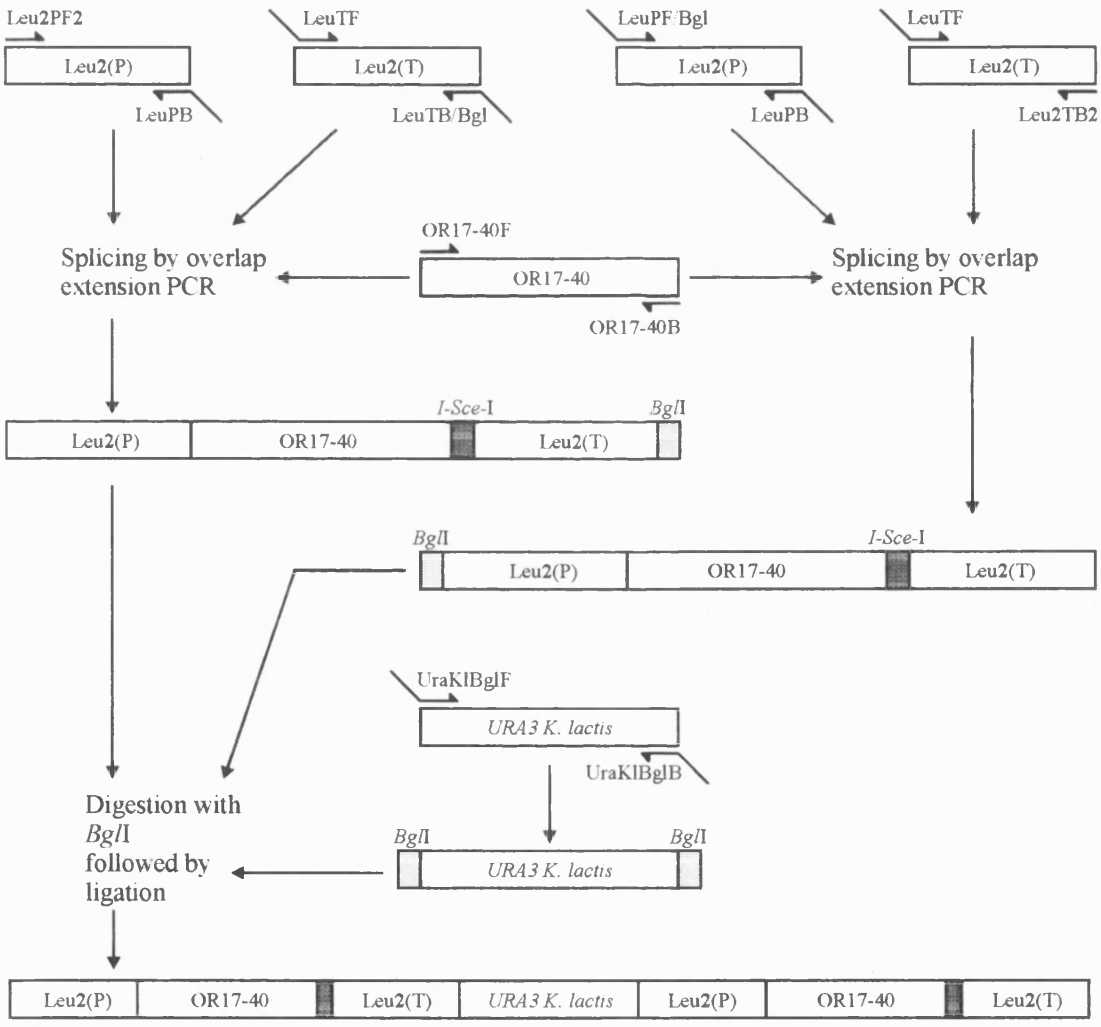


Figure 3.9 Schematic representation of the synthesis of the disruption cassette for the integration of the *OR17-40* ORF at the *LEU2* locus. The extreme flanking sequences in the completed disruption cassette, Leu2(P)/Leu2(T), contain at their ends, sequences not present in the internal repeats of these structures, allowing the cassette to be amplified by PCR.

The reactions performed in the construction of the disruption cassette are described in detail next. The overall scheme showing the various DNA fragments, primer binding sites, and the steps involved in the construction of the completed disruption cassette, is shown in Figure 3.9.

A) *Primers to amplify the LEU2 promoter (P) and terminator (T) regions*

The sequences of the primers (Table 3.1 and Table 5.3 [Leu2PF2/LEU2TB2]) used to generate the flanking regions in the disruption cassette were derived (as discussed above) from the promoter and terminator regions of the *LEU2* locus. The DNA sequences were obtained from the *Saccharomyces cerevisiae* genome database (Section 2.2.1). The accession number for the *LEU2* open reading frame is YCL018W.

The LeuPB and LeuTF primers were designed to have their 5'-ends homologous to 24 nucleotides at either end of the *OR17-40* ORF. This was to allow the *LEU2* flanking sequences, once amplified, to be joined to the odorant receptor fragment via a 'splicing by overlap extension' PCR (Horton *et al.*, 1989).

The LeuTB/Bgl and LeuPF/Bgl primers incorporated *Bgl*I sites in their 5' ends to allow the direct repeats, once amplified, to be ligated to the *K. lactis URA3* gene (also flanked by *Bgl*I sites).

B) *PCR amplification of the LEU2 promoter and terminator regions*

Four separate Expand™ PCRs (Section 2.2.3b) were performed to amplify the *LEU2* fragments (top of Figure 3.9) from yeast genomic DNA. The genomic DNA was prepared from the *S. cerevisiae* **a**-strain FY3 (Section 3.2.1) following the procedure outlined in Section 2.2.13.

The primer combinations Leu2PF2/LeuPB and LeuPF/Bgl/LeuPB were used to generate the promoter region fragments. For both primer combinations the reaction parameters were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s + 20 s/cycle) × 15; 72°C, 5 min].

The primer combinations LeuTF/LeuTB/Bgl and LeuTF/Leu2TB2 were used to generate the terminator region fragments. For both primer combinations the reaction parameters were as for the promoter regions, above.

All PCR products were purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c). The purified fragments were quantified by agarose gel electrophoresis (Section 2.2.4).

C) *Amplification the OR17-40 ORF*

The primer pair OR17-40F/OR17-40B was used to amplify the *OR17-40* ORF from plasmid pYES2[*OR17-40*] (Section 3.2.2a). The sequences of the primers (Table 3.1) were derived form the start and end of the *OR17-40* open reading frame (Ben-Arie *et al.*, 1993).

An Expand™ PCR (Section 2.2.3b) was performed using the following parameters: [94°C, 2 min; (94°C, 15 s; 58°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 58°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

The PCR product was purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c) then quantified by agarose gel electrophoresis (Section 2.2.4).

D) *PCR mediated generation of the repeat sequences*

Two separate Expand™ PCRs (Section 2.2.3b) were set up to amplify the left hand side (LHS) and right hand side (RHS) direct repeat sequences. Approximately equal amounts (~40 ng) of the appropriate *LEU2*-promoter fragment and *LEU2*-terminator fragment (as shown for both splicing by overlap extension PCRs, Figure 3.9) and the *OR17-40* fragment, were added together with either the primer pair Leu2PF2/LeuTB/Bgl, or LeuPF/Bgl/Leu2TB2. The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 30 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 30 s + 20 s/cycle) × 15; 72°C, 5 min].

The assembled PCR products were purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c), and the fragments quantified by agarose gel electrophoresis (Section 2.2.4).

E) *Primers to amplify the URA3 gene from K. lactis*

The primers UraKlBglF/UraKlBglB (Table 3.1) were designed to amplify the *K. lactis URA3* gene - the same fragment as used by Längle-Rouault and Jacobs (1995) - from vector pTG5756 (Section 3.2.2c). The *URA3* DNA sequence

was obtained from the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov). The accession number for the sequence is D00431.

Both primers contain a *BglI* restriction endonuclease site in their 5'-ends to allow subsequent ligation of the amplified fragment to the direct repeat sequences.

F) *PCR amplification of the URA3 gene from K. lactis*

An Expand™ PCR (Section 2.2.3b) using the primers UraklBglF and UraKlBglB was performed to amplify the *K. lactis URA3* gene from vector pTG5756 (Section 3.2.2c). The reaction was performed for 25 cycles using the following parameters: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

The amplified fragment was purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c), and quantified by agarose gel electrophoresis (Section 2.2.4).

G) *Generation of the completed disruption cassette*

Approximately equal amounts (~0.3 μg) of the purified LHS direct repeat fragment, RHS direct repeat fragment and the *K. lactis URA3* fragment were separately digested. Each was digested in a total digest volume of 20 μl, with *BglI* (20 units) at 37°C overnight, then purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c), and quantified (Section 2.2.4).

Approximately equal amounts (~0.2 μg) of the cleaned digested fragments were then joined by ligation (Section 2.2.15). The ligation product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

H) *PCR Amplification of the completed disruption cassette*

In order to generate enough of the completed disruption cassette for transformation purposes the purified cassette was amplified via an Expand™ PCR (Section 2.2.3b) using the primers Leu2PF2 and Leu2TB2 (Table 5.3). The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 68°C, 4 min) × 10; (94°C, 15 s; 55°C, 30 s; 68°C, 4 min + 20 s/cycle) × 15; 68°C, 5 min].

The amplified PCR product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

Primer name	Sequence (5' - 3')	Description (5' - 3')
LeuPB	ATTGGYCCCAGMTTCTGGCTSCATTA GAATGGTATATCCTTGAA	24 n.t. homology to template strand at the start of the *OR17-40* ORF; 20 n.t. homology to the *LEU2* template strand immediately before the initiation codon.
LeuTF	GGGAGGMGRTCACTGRCYTGATAAGT TACGC**TAGGGATAACAGGGTAAT**AC TGTCGCCGAAGAAGTTAAG	24 n.t. homology (including an added TAA stop codon) to end of *OR17-40* ORF (coding strand); 5 n.t. spacer; an I-*Sce*I site; a 1 n.t. spacer; 20 n.t. homology to coding strand of *LEU2*, 35 - 15 b.p. upstream from the termination codon.
LeuTB/BgI	CAGAT**GCCACTCAGGC**AAATTAGGA ATCATAGTTTC	5 n.t. overhang; a *BglI* site; 20 n.t. homology to template strand of *LEU2*, 285 - 305 b.p. downstream from the termination codon.
LeuPF/BgI	CAGAT**GCCTCCATGGC**ATCAAATTC GATGACTGGAA	5 n.t. overhang; a *BglI* site; 20 n.t. homology to coding strand of *LEU2*, 300 - 280 b.p. upstream from initiation codon.
OR17-40F	ATGCAGCCAGAATCTGGGGC	Binds to the first 20 n.t. of the template strand of the *OR17-40* ORF.
OR17-40B	TCAAGCCAGTGACCGCCTCC	Binds to the final 20 n.t. of the coding strand of the *OR17-40* ORF.
UraKlBglF	CAGAT**GCCTGAGTGGC**GTGATTCTG GGTAGAAGA	5 n.t. overhang; a *BglI* site; 18 n.t. homology to the coding strand of the *K. lactis URA3* gene, 333 - 315 b.p. upstream from the initiation codon.
UraKlBglB	CAGAT**GCCATGGAGGC**AGTTGATCC ATTGTGTGC	5 n.t. overhang; a *BglI* site; 18 n.t. homology to the template strand of the *K. lactis URA3* gene, 333 - 315 b.p. downstream from the termination codon.

Table 3.1 Sequences of the oligonucleotides used in the construction of the cassette for the integration of *OR17-40* at the *LEU2* locus. Restriction endonuclease sites are in bold. Degenerate bases are as follows: M: A or C, K: G or T, S: G or C, Y: C or T, and R: A or G.

3.3.2 Integration of the *OR17-40* ORF at the *LEU2* locus of *Saccharomyces cerevisiae* strain JRY50

A) *Transformation of the disruption cassette into JRY50*

The purified disruption cassette (~0.1 µg), was transformed into the yeast **a**-strain JRY50 (Section 3.2.1) using a lithium acetate/polyethyleneglycol transformation protocol (Section 2.2.10b). Transformants were inoculated onto SC-Ura plates (Section 2.2.7a) and incubated at 30°C for 3 days.

B) *Verification of correct cassette integration by PCR*

Six Ura^+ colonies were selected from which glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Both novel joints (Figure 3.10) formed upon integration were screened via *Taq* PCRs (Section 2.2.3a). Correct integration of the LHS of the cassette into the genome was checked using the primer pair Leu2PF3[Table 5.4]/Mable(B)[Table 3.2], and correct integration of the RHS was checked using the primer pair Leu2TB3[Table 5.4]/Fred(F)[Table 3.2]. For both reactions the cycling parameters were: [94°C, 2 min; (94°C, 30 s; 40°C, 1 min; 72°C, 1 min 40 s) × 35].

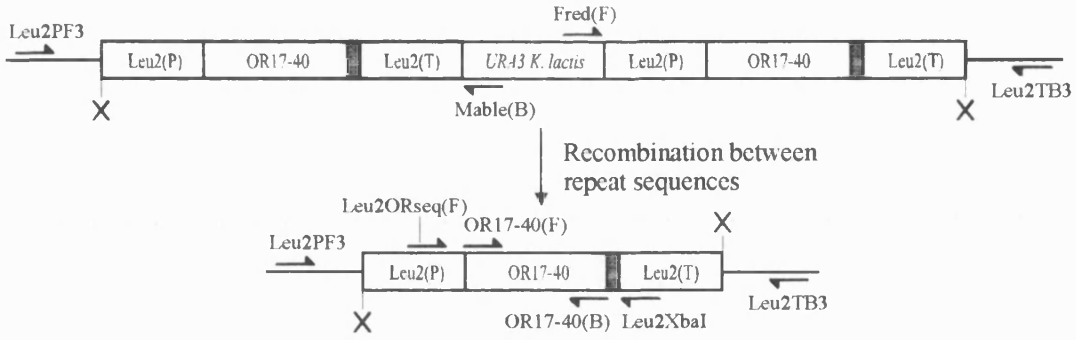


Figure 3.10 Position and orientation of the primers used to verify correct integration of the disruption cassette at the *Leu2* locus. Symbol '×' represents the junction between the introduced DNA fragment and the chromosomal DNA (thin line). Shaded box: I-*Sce*I site (Figure is not to scale).

C) *Selection and screening for cassette recombination: loss of URA3 marker*

A single positive colony (from the above transformation) was grown overnight at 30°C in YPD (10 ml). The cells were pelleted by centrifugation (2000 r.p.m., 5 min) and resuspended in water (10 ml). Approximately 2.5×10^6 washed cells were spread evenly onto 5-FOA selective medium (Section 2.2.7b) and incubated for 5 days at

30°C. After this period eight large distinct colonies were picked and restreaked (in order to avoid contamination from micro-colony background) onto fresh 5-FOA medium, and allowed to grow for 2 days. Individual colonies were picked into YPD (10 ml) and grown overnight at 30°C. Glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13) from the resulting cultures.

Correct loss of the *URA3* marker following mitotic recombination was determined by *Taq* PCR screening (Section 2.2.3a). A 'straight through' screen using the primers Leu2PF3/Leu2TB3 was performed (Figure 3.10), the cycling parameters were: [94°C, 2 min; (94°C, 30 s; 42°C, 1 min; 72°C, 1 min 30 s) × 35]. Using the primers Leu2PF3/OR17-40(B) and Leu2TB3/OR17-40(F)[Figure 3.10] the novel junctions created following marker excision were examined, for both reactions the cycling parameters were as above.

Primer name	**Sequence (5' - 3')**	**Description (5' - 3')**
Fred(F)	AGACTGTTTGTCATCCAA	Binds to the template strand of the *K. lactis URA* gene, 248 - 266 b.p. downstream from the termination codon.
Mable(B)	GCTCGAGGCAAGTAATGT	Binds to the coding strand of the *K. lactis URA* gene, 210 - 218 b.p. upstream from the initiation codon.
Leu2ORseq(F)	GTGGTTAGCAATCGTCTTAC	20 n.t. homology to coding strand of *LEU2*, 89 - 69 b.p. upstream from initiation codon.

Table 3.2 Sequences of the oligonucleotide primers used for screening and sequencing of the disruption cassette at the *LEU2* locus in JRY50.

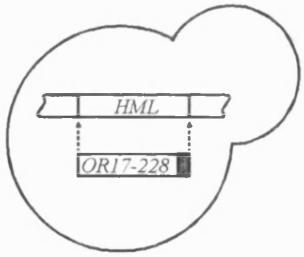
D) *Verifying the I-SceI site in the integrated fragment*

By DNA sequencing: An Expand™ PCR (Section 2.2.3b) was performed using the primer pair Leu2PF3/Leu2TB3 to amplify the *LEU2* locus from JRY50 genomic DNA (from above). The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 20 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 20 s + 20 s/cycle) × 15;

72°C, 5 min]. The DNA fragment was analysed (Section 2.2.4) and then purified by LMP-agarose gel electrophoresis (Section 2.2.4a). The sample was recovered from the gel by β-agarase digestion (Section 2.2.4b). Using the purified DNA fragment three sequencing reactions were performed (Section 2.2.16) using the primers Leu2ORseq(F) [Table 3.2], Leu2XbaI (Table 5.5) and Leu2PF2 (Table 5.3). The relative positions of the sequencing primers is shown in Figure 3.10 (the binding site of Leu2PF2 has been shown previously, Figure 3.9).

By restriction digest: Approximately 80 ng of the above purified DNA fragment was digested in a total volume of 20 µl with I-*Sce*I (5 units) at 37°C for 1 hour. Digested DNA (3 µl) was analysed by agarose gel electrophoresis (Section 2.2.4).

Overview: The next section describes the steps performed in order to integrate the *OR17-228* ORF (followed by the I-*Sce*I recognition site) at the *HML* locus in the experimental strain under construction.



3.3.3 Strategy for the integration of the *OR17-228* ORF at the donor locus *HML*

Two different gene targeting strategies were attempted in order to integrate the *OR17-228* ORF at the donor locus, *HML*. A direct gene replacement strategy (Boeke *et al.*, 1984; Rothstein, 1991; Wach, 1996) and a two-step 'pop-in/pop-out' strategy (Scherer and Davis, 1979). The original strategy adopted, the direct gene replacement method, was to prove unsuccessful. However, its application is described here (in addition to the 'pop-in pop-out' strategy), as it is of interest, and relevance, to the construction of the experimental system, particularly to the choice of yeast strain JRY50 (Section 3.2.1).

The direct gene replacement strategy was simpler overall (involving the construction of less cumbersome DNA fragments) than the method used to target the *OR17-40* ORF to the *LEU2* locus (Section 3.3.1). The strategy was still based upon counter selection of a *URA3* marker (Section 3.3.1), except that marker loss in this case was a result of a second transformation reaction and not mitotic recombination. Thus, a *URA3* gene is first integrated within the locus to be mutagenized, and then the recipient strain transformed with linear, mutated DNA, spanning the point at which *URA3* is inserted, followed by 5-FOA selection. In order for the strategy to work it is of course necessary that the *URA3* gene be transcriptionally active; for the initial selection of Ura^+ cells, and the subsequent counterselection step. As described in Section 1.1.1, the *HML* locus is maintained in a transcriptionally silenced state due to the action of the *SIR* proteins and other *trans*-acting factors at the *E* and *I* regulatory elements. Yeast strain JRY50 was chosen as it contains a temperature sensitive mutation in the gene *SIR4* (Schnell and Rine, 1986), whereby silencing is not maintained at the restrictive temperature (34°C), allowing constitutive expression (of *URA3*) from the *HML* locus. Later, repression at the *HML* locus must be re-established to prevent the endonuclease I-*Sce*I, when induced, from cutting the donor gene (Section 1.3). This is achieved by growing JRY50 at the permissive temperature (24°C). The steps involved in the construction of the necessary DNA fragments for the direct gene targeting strategy, i.e. a *K. lactis-URA3* containing cassette, and a 'mutated-DNA' cassette, are shown in Figure 3.11, and described in detail in subsequent sections.

The second integration strategy, the 'pop-in/pop-out' replacement method, is similar to the method used to target the *OR17-40* ORF to the *LEU2* locus, involving homologous recombination between direct repeats. The desired mutation (the *OR17-228* ORF, followed by an I-*Sce*I restriction site) is flanked by regions of homology to the desired target, the *HML* locus. This fragment is then cloned into an integrating vector that contains the *URA3* selectable marker. Integration of the circular molecule results in direct repeats of the homologous sequences, with one mutant copy and one wild-type copy (Figure 3.13). Pairing of the homologous direct repeats and recombination results in loss of the plasmid sequences. As before, this event is selected on 5-FOA medium. Cross-overs that occur on the appropriate side of the mutant site replace the wild-type chromosomal site with the mutant sequence. The

fragment used for cloning into the integrating vector was the one made for the 'direct replacement' strategy (Figure 3.11b). The fragment was modified to include appropriate restriction sites at its 5'- and 3'-ends for ease of cloning. The steps involved are described in detail in the appropriate sections.

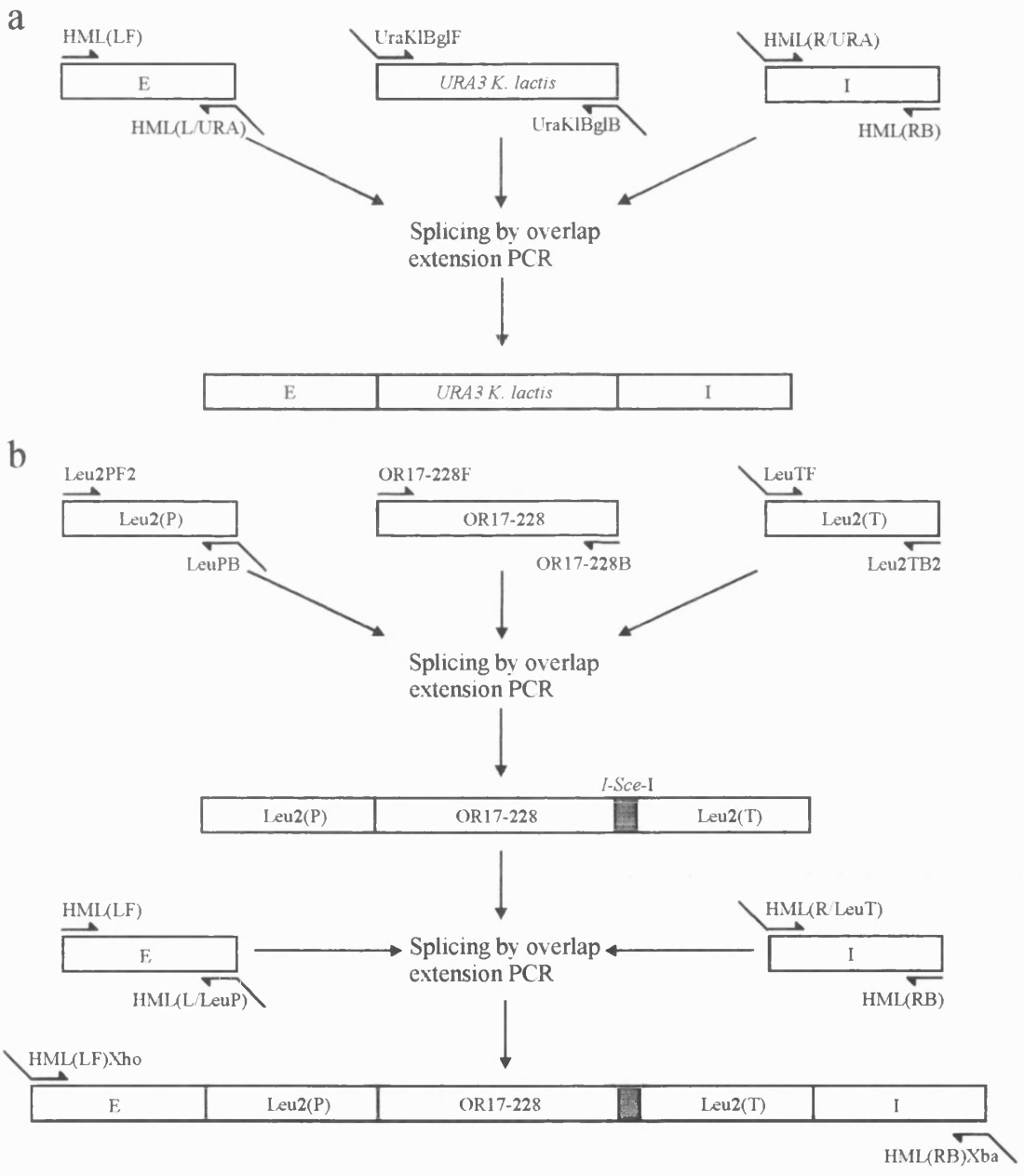


Figure 3.11 **a**. Schematic representation of the synthesis of a cassette for the integration of the *K. lactis URA3* gene at the *HML* locus. **b**. The steps involved in the synthesis of a cassette for the introduction of the *OR17-228 ORF* at the *HML* locus. Note that the *OR17-228* ORF is immediately flanked by regions of homology to the acceptor locus, i.e. the Leu2(P) and Leu2(T) fragments, for the reasons described in Section 1.3. The primers shown on the completed cassette, HML(LF)Xho and HML(RB)Xba, were those used to amplify the fragment for subsequent cloning into the integrating vector (Figure is not to scale).

3.3.4 Construction of the *E-URA3-I HML* disruption cassette

A) *Primers to amplify flanking segments from the E and I regions of the HML locus*

The sequences of the primers (Table 3.3) used to generate the flanking segments in the disruption cassette were derived from the *E* and *I* regions of the *HML* locus. The DNA sequences were obtained from the *Saccharomyces cerevisiae* Genome Database (Section 2.2.1) coordinates 10782 to 16309 (corresponding to *HMLALPHA-I* ± 2500 n.t.). Primer binding sites are described with respect to the *ALPHA1* ORF.

The HML(L/URA) and HML(R/URA) primers were designed to have their 5'-ends homologous to the *Bgl*I 'tails' of the primers UraKlBglF and UraKlBglB, respectively. This was to allow the *E* and *I* sequences, once amplified, to be joined to the odorant receptor fragment via a 'splicing by overlap extension' PCR.

Primer name	Sequence (5' - 3')	Description (5' - 3')
HML(LF)	TAGAAGACAAGTAGCGCAGT	Binds to the template strand, 2212 - 2192 b.p. upstream from *α1* initiation codon.
HML(RB)	GACCTCAACTTAGTAAAATG	Binds to the coding strand, 1152 - 1132 b.p. downstream from *α1* termination codon.
HML(L/URA)	AGAATCAC**GCCACTCAGGC**A TCTGTTTCATTCTATGTGCGCT AG	24 n.t. homology to the 5'-end of primer UraKlBglF; 20 n.t. homology, 1828 - 1848 b.p. upstream of *α1* initiation codon.
HML(R/URA)	GATCAACT**GCCTCCATGGC**A TCTGTGAGTAACTAACTCTCA TGG	24 n.t. homology to the 5'-end of primer UraKlBglB; 20 n.t. homology, 646 - 666 b.p. downstream of *α1* termination codon.

Table 3.3 Sequences of the oligonucleotides used in the construction of the *E-URA3-I HML* disruption cassette. Restriction endonuclease sites are indicated in bold.

B) *PCR amplification of the E and I regions*

Two separate Expand™ PCRs (Section 2.2.3b) were performed to amplify the *E* and *I* DNA fragments (top of Figure 3.11a) from yeast genomic DNA. The genomic DNA was prepared from JRY50 (Section 3.2.1) following the procedure outlined in Section 2.2.13.

The primer pair HML(LF)/HML(L/URA) was used to generate the *E* DNA fragment, and the primer pair HML(R/URA)/HML(RB) to generate the *I* DNA fragment. For both reactions the cycling parameters were: [94°C, 2 min; (94°C, 15 s; 52°C, 30 s; 72°C, 45 s) × 10; (94°C, 15 s; 52°C, 30 s; 72°C, 45 s + 20 s/cycle) × 15; 72°C, 5 min].

The PCR products were purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c), and quantified by agarose gel electrophoresis (Section 2.2.4).

C) *PCR mediated generation of the E-URA3-I HML disruption cassette*

Approximately equal amounts (~15 ng) of the *K. lactis URA3* DNA fragment (Section 3.3.1f), the *E* DNA fragment and the *I* DNA fragment were added together with the primers HML(LF) and HML(RB) [Table 3.3] in an Expand™ PCR. The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 48°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 48°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min].

The amplified PCR product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

3.3.5 Integration of the *E-URA3-I HML* disruption cassette

A) *Transformation of the disruption cassette into strain BGY2*

The purified *E-URA3-I HML* disruption cassette (~0.4 μg) was transformed by electroporation (Section 2.2.10d) into strain BGY2 (Section 3.2.1). Transformants were inoculated onto SC-Ura plates (Section 2.2.7a) and incubated at 30°C for 3 days.

B) *Verification of correct cassette integration by PCR*

Twelve Ura^+ colonies were selected from which glycerol stock (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Both novel joints (Figure 3.12) formed upon integration were screened via *Taq* PCRs (Section 2.2.3a). Correct

integration of the LHS of the cassette into the genome was checked using the primer pair SHML(LF)[Table 3.4]/Mable(B)[Table 3.2], and correct integration of the RHS was checked using the primer pair SHML(RB)/UraKLSF (Table 3.4). For both reactions the cycling parameters were: [94°C, 2 min; (94°C, 30 s; 38°C, 1 min; 72°C, 45 s) × 35]

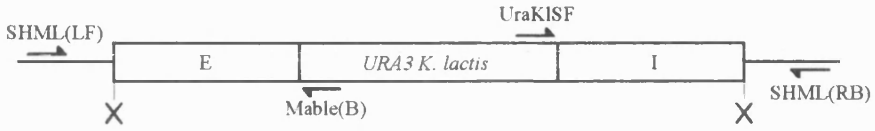


Figure 3.12 Position and orientation of the primers used to verify correct integration of the *E-URA3-I* disruption cassette at the *HML* locus. Symbol '×' represents the junction between the introduced DNA fragment and the chromosomal DNA (thin line). Figure is not to scale.

Primer name	Sequence (5' - 3')	Description (5' - 3')
UraKlSF	CCAGACTGTTTGTCATCCAA	Binds to the template strand of the *K. lactis URA3* gene, 266 - 246 b.p. downstream from the termination codon.
SHML(LF)	TAAGAACGTATTTTGGATGG	Binds to template strand, 2329 - 2308 b.p. upstream from *α1* initiation codon.
SHML(RB)	TTACGTAGAATTTGACACAA	Binds to coding strand, 1222 - 1202 b.p. downstream from *α1* termination codon.

Table 3.4 Sequences of the oligonucleotide primers used to screen for correct cassette integration at the *HML* locus.

3.3.6 Construction of the *E-OR17-228-I* gene replacement cassette

A) *Amplification of the OR17-228 ORF*

The primer pair OR17-228F/OR17-228B (Table 3.5) was used to amplify the *OR17-228* ORF from plasmid pYES2[*OR17-228*] (Section 3.2.2f). The sequences of

the primers (Table 3.5) were derived form the start and end of the *OR17-228* open reading frame (Ben-Arie *et al.*, 1993).

An Expand™ PCR (Section 2.2.3b) was performed using the following parameters: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s + 20 s/cycle) × 15; 72°C, 5 min].

The PCR product was purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c), then quantified by agarose gel electrophoresis (Section 2.2.4).

Primer name	**Sequence (5' - 3')**	**Description (5' - 3')**
OR17-228(B)	TCAGGTCAGTGATCTCCTCC	Binds to the final 20 n.t. of the coding strand of the *OR17-228* open reading frame.
OR17-228(F)	ATGGAGCCAGAAGCTGGGAC	Binds to the first 20 n.t. of the template strand of the *OR17-228* open reading frame.
HML(R/LeuT)	ATTTATATATTGGAGGATTTT CTCTGAGTAACTAACTCTCAT GG	24 n.t. homology to the 3'-end of the *LEU2*-terminator fragment (as with primer Leu2TB2); 20 n.t. homology, 646 - 666 b.p. downstream of *α1* termination codon.
HML(L/LeuP)	TGATTCTGTGCGATAGCGCCC CTGTTTCATTCTATGTGCGCTA G	24 n.t. homology to the 5'-end of the *LEU2*-promoter fragment (amplified with primer Leu2PF2); 20 n.t. homology, 1828 - 1848 b.p. upstream of *α1* initiation codon.

Table 3.5 Sequences of the oligonucleotide primers used in the construction of the *E-OR17-228-I* gene replacement cassette.

B) *Generation of the 'central portion' of the gene replacement cassette*

Approximately equal amounts (~40 ng) of the *LEU2*-promoter fragment (amplified with the primers Leu2PF2/LeuPB [Section 3.3.1b]) and *LEU2*-terminator fragment

(amplified with the primers LeuTF/Leu2TB2 [Section 3.3.1b]) and the *OR17-228* fragment (above Section), were added together in an Expand™ PCR (Section 2.2.3b) with the primers Leu2PF2 and Leu2TB2. The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min, 30 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 30 s + 20 s/cycle) × 15; 72°C, 5 min].

The assembled PCR product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

C) *PCR amplification of the E and I regions*

The *E* and *I* regions of the gene replacement cassette were constructed as described in Section 3.3.4b except that the primers HML(L/URA) and HML(R/URA) were replaced with the primers HML(L/LeuP) and HML(R/LeuT) respectively. The new primers (Table 3.5) were designed with their 5'-ends homologous to the 5'-end of the *LEU2*-promoter fragment [primer HML(L/LeuP)] and the 3'-end of the *LEU2* terminator fragment [primer HML(R/LeuT)], enabling the *E* and *I* fragments, once amplified, to be joined via a 'splicing by overlap extension' PCR to the (above) 'central portion'.

D) *PCR mediated generation of the E-OR17-228-I gene replacement cassette*

Approximately equal amounts (~20 ng) of the *E* DNA fragment, the *I* DNA fragment and the 'central portion' (Section 3.3.6b) were added together with the primers HML(LF) and HML(RB) (Table 3.3). An Expand™ PCR (Section 2.2.3b) was performed with the following cycling parameters: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min].

The assembled PCR product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

3.3.7 Integration of the *E-OR17-228-I* gene replacement cassette

A) *Transformation of the gene replacement cassette into BGY3*

The purified disruption cassette (~0.4 µg), was transformed into BGY3 (Section 3.2.1) using a lithium acetate/polyethyleneglycol transformation protocol (Section 2.2.10b).

The final incubation step was increased from 2 h to 4 h in order to allow expression of the Ura⁻ phenotype. The cells were then inoculating onto 5-FOA plates (Section 2.2.7b) and incubated at 30°C for 5 days.

B) *PCR screening of transformants for integration of the gene replacement cassette*
Gene replacement events at the *HML* locus were analysed by means of a 'straight-through' PCR screen (Section 2.2.12b). An Expand™ PCR was performed using the primers SHML(LF) and SHML(RB) [Table 3.4], the cycling parameters for the reaction were: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 68°C, 3 min) × 10; (94°C, 15 s; 55°C, 30 s; 68°C, 3 min + 20 s/cycle) × 15; 68°C, 5 min].

3.3.8 Construction of integrating vector pRS306[*OR17-228*]

A) *Primers to allow cloning of the E-OR17-228-I gene replacement cassette*
In order that the previously made *E-OR17-228-I* gene replacement cassette (Section 3.3.6d) could be cloned into the integrating vector pRS306 (Section 3.2.2d) two new primers were designed. The primers were the same as those used in the construction of the replacement cassette, HML(LF) and HML(RB), except that the new primers incorporate in their 5'-ends either a *Xho*I or *Xba*I restriction endonuclease site (Figure 3.11b). The sequences of the primers is given in Table 3.6.

Primer name	Sequence (5' - 3')	Description (5' - 3')
HML(LF)Xho	CAGAT**CTCGAG**TAGAAGACAAG TAGCGCAGT	5 n.t. spacer; a *Xho*I site; binds to the template strand, 2212 - 2192 b.p. upstream from *α1* initiation codon.
HML(RB)Xba	CAGAT**TCTAGA**GACCTCAACTTA GTAAAATG	5 n.t. spacer; a *Xba*I site; binds to the coding strand, 1152 - 1132 b.p. downstream from *α1* termination codon.

Table 3.6 Sequences of the oligonucleotide primers used to clone the *E-OR17-228-I* gene replacement cassette into the integrating vector pRS306. Restriction endonuclease sites are shown in bold.

B) *PCR amplification of the E-OR17-228-I gene replacement cassette for cloning*

An Expand™ PCR (Section 2.2.3b) was performed using the *E-OR17-228-I* gene replacement cassette (Section 3.3.6d) and the primers HML(LF)Xho and HML(RB)Xba. The cycling parameters were: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 55°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min].

The amplified product was purified by LMP-agarose gel electrophoresis (Section 2.2.4b) and analysed (Section 2.2.4).

C) *Cloning of the gene replacement cassette into pRS306*

Plasmid pRS306 (~0.8 µg)[Section 3.2.2d] was double-digested (Section 2.2.14) with *Xho*I (30 units) and *Xba*I (30 units) at 37°C overnight. The cut-vector was purified by LMP-agarose gel electrophoresis (Section 2.2.4a) and recovered by β-agarase digestion (Section 2.2.4b). The *E-OR17-228-I* gene replacement cassette (from above) was digested and purified in the same manner. Both double-digested products were quantified by agarose gel electrophoresis (Section 2.2.4). The gene replacement cassette (~40 ng) was then ligated into the vector (~20 ng)[Section 2.2.15] and transformed directly by electroporation into competent *E. coli* (Section 2.2.9d). Thirteen ampicillin resistant colonies were PCR screened (Section 2.2.12a) for the presence of the cloned fragment using primers T7 and M13 (Table 5.2). Positive transformants were grown overnight in 4 ml LB (Section 2.2.6) from which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

3.3.9 Integration of vector pRS306[*OR17-228*] into yeast strain BGY2

In order to ensure that a transformed plasmid integrates at the desired chromosomal location, the integration event is targeted by the introduction of a double-strand break in the plasmid (Orr-Weaver *et al.*, 1981). When the double-strand break is made within the yeast sequence on a plasmid without an *ARS* sequence, the frequency of transformation increases 10- to 1000-fold compared to uncut plasmid. The double-strand break directs plasmid integration to the chromosomal region homologous to the cut sequence.

Prior to transformation, for the reasons given above, the vector (~0.5 μg) was digested overnight (Section 2.2.14) with the restriction enzyme *Eco*RI (40 units). The enzyme makes a single cut in the *E* region of the cloned DNA fragment (and nowhere else in the vector) to produce free 5'- and 3'-ends that have 168 b.p. and 235 b.p. homology to the target locus respectively. Successful digestion was checked by agarose gel electrophoresis (Section 2.2.4). The cut vector was purified using the Wizard™ DNA Clean-Up System (Section 2.2.4c).

The linearized vector (~80 ng) was transformed by electroporation (Section 2.2.10d) into strain BGY2 (Section 3.2.1). The transformed cells were inoculated onto SC-Ura plates (Section 2.2.7a) and incubated at 30°C for 3 days.

A) *Verification of correct integration of plasmid pRS306[OR17-228]*

Six Ura^+ colonies were selected from which glycerol stock (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Correct integration of the vector at the *HML* locus was first checked by means of a 'straight-through' PCR screen using the primers SHML(LF) and SHML(RB) [Table 3.4]. An Expand™ PCR (Section 2.2.3b) was performed, the cycling parameters for the reaction were: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 68°C, 6 min 30 s) × 10; (94°C, 15 s; 55°C, 30 s; 68°C, 6 min 30 s + 20 s/cycle) × 15; 68°C, 10 min].

In addition to the above screen, both novel joints (Figure 3.13) formed upon integration were also verified via Expand™ PCRs. Correct integration of the LHS of the vector into the genome was checked using the primer pair SHML(LF)/T7 [Table 5.2], the cycling parameters were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min]. Correct integration of the RHS was checked using the primer pair SHML(RB)/M13 (Table 5.2), the cycling parameters were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 3 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 3 min + 20 s/cycle) × 15; 72°C, 5 min].

All PCR screening products were analysed by agarose gel electrophoresis (Section 2.2.4).

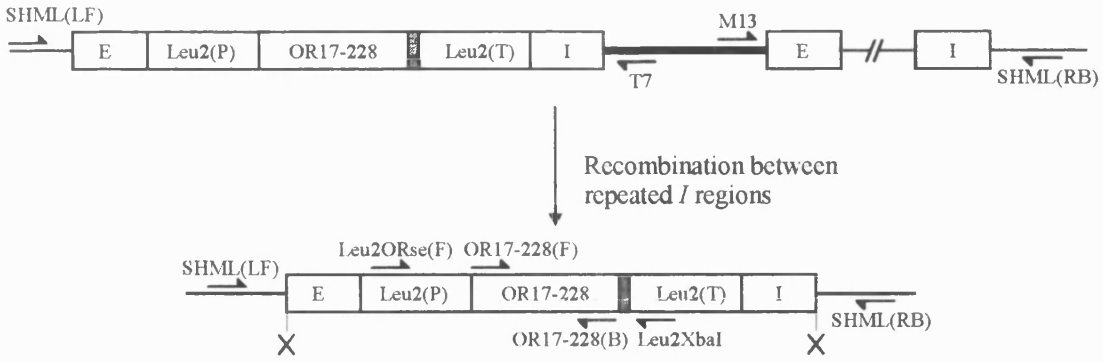


Figure 3.13 Position and orientation of the primers used to verify correct integration (and subsequent loss) of plasmid pRS306[*OR17-228*] at the *HML* locus. Symbol '×' represents the junction between the introduced DNA fragment and the chromosomal DNA (thin line). Shaded box: I-*Sce*I site. pRS306 DNA: thick line. Note that excision of the integrated DNA can proceed via two routes: recombination between the repeated *E* regions, resulting in the wild-type *HML* locus (not shown), and recombination between the repeated *I* regions (as shown) resulting in integration of the *OR17-228* ORF at the *HML* locus (Figure is not to scale).

B) *Selection and screening for vector recombination: loss of URA3 marker*

A single positive colony (from the above transformation) was grown overnight at 30°C in SC-Ura (10ml). A loopfull of cells was streaked onto 5-FOA selective medium (Section 2.2.7b) and incubated for 5 days at 30°C. After this period eight large distinct colonies were picked and restreaked (in order to avoid contamination from micro-colony background) onto fresh 5-FOA medium, and allowed to grow for 2 days. Individual colonies were picked into YPD (10 ml) and grown overnight at 30°C. Glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13) from the resulting cultures.

Excision of vector sequences from the picked colonies was first confirmed by means of a 'straight through' PCR screen. The primers and the reaction conditions were as used initially for the verification of correct integration of pRS306[*OR17-228*] (Section 3.3.9a).

A single colony showing excision of the vector to leave the transgene at the target locus was further characterised by PCR screening using Expand™ (Section 2.2.3b). Both novel junctions formed upon vector excision were examined (Figure 3.13). The LHS junction was examined using the primer pair SHML(LF)[Table 3.4]/OR17-228(B)[Table 3.5], and the RHS junction was examined using the primer pair SHML(RB)[Table 3.4]/OR17-228(F)[Table 3.5]. The cycling

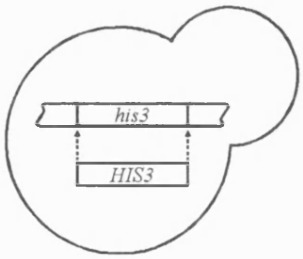
conditions for both reactions were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min].

C) *Verifying the I-SceI site in the integrated fragment*

By DNA sequencing: An Expand™ PCR (Section 2.2.3b) was performed using the primer pair SHML(LF)/SHML(RB) to amplify the *HML* locus from BGY2 genomic DNA (from above). The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 20 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min 20 s + 20 s/cycle) × 15; 72°C, 5 min]. The DNA fragment was analysed (Section 2.2.4) and then purified by LMP-agarose gel electrophoresis (Section 2.2.4a). The sample was recovered from the gel by β-agarase digestion (Section 2.2.4b). Using the purified DNA fragment three sequencing reactions were performed (Section 2.2.16) using the primers Leu2ORseq(F) [Table 3.2], Leu2XbaI (Table 5.5) and Leu2PF2 (Table 5.3). The relative positions of the sequencing primers is shown in Figure 3.13 (the binding site of Leu2PF2 has been shown previously, Figure 3.9).

By restriction digest: Approximately 80 ng of the above purified DNA fragment was digested in a total volume of 20 µl with I-*Sce*I (5 units) at 37°C for 1 hour. Digested DNA (3 µl) was analysed by agarose gel electrophoresis (Section 2.2.4).

Overview: The next section describes the conversion of the experimental strain under construction to a His^+ phenotype.



3.3.10 Conversion of the experimental strain to a His^+ phenotype

As described in Section 1.3, the *HIS4* gene, located between the donor and acceptor loci, is to be used as a selective marker against undesired recombination events following induction of gene conversion. For this strategy to work it is of course a

requirement that the experimental strain be His^+. JRY50, the parent strain, is His^- due to a mutation in the gene *HIS3*. In order that the *HIS4* maker could still be used as intended the below work was carried out to render the experimental strain His^+.

A) *Primers to amplify the HIS3 locus*

The DNA sequence of the *HIS3* ORF (including upstream and downstream sequences) was obtained from the *Saccharomyces cerevisiae* genome database (Section 2.2.1), coordinates 721944 - 722606 on chromosome XV. The primers to amplify the *HIS3* locus (Table 3.7) were designed with reference to this sequence. The primers were designed to amplify the *HIS3* ORF and a proportion of upstream and downstream sequences. Approximately 1 k.b.p. of promoter sequence was incorporated into the amplified fragment, and ~300 b.p. of terminator sequence. The His3TB primer was positioned so as not to encroach on the nearby ORF of the *DED1* gene.

B) *PCR amplification of the HIS3 locus*

Using the primer pair HIS3PF/HIS3TB and genomic DNA from the His^+ strain FY10 (Section 5.2.1), an Expand™ PCR was performed. The cycling parameters for the reaction were: [94°C, 2 min; (94°C, 15 s; 39°C, 30 s; 72°C, 1 min, 30 s) × 10; (94°C, 15 s; 39°C, 30 s; 72°C, 1 min 30 s + 20 s/cycle) × 15; 72°C, 5 min].

The amplified fragment was analysed by agarose gel electrophoresis (Section 2.2.4) and purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c).

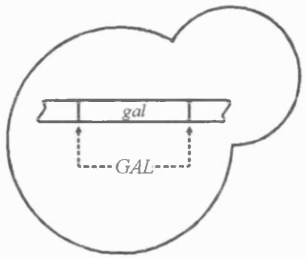
Primer name	**Sequence (5' - 3')**	**Description (5' - 3')**
HIS3PF	GTTATTTCATCCAGATATAA	Bind to the template strand, 986 - 966 b.p. upstream from the *HIS3* initiation codon.
HIS3TB	TTCTTGTTGTTCTTACGGAA	Binds to coding strand, 289 - 269 b.p. downstream from the *HIS3* termination codon.

Table 3.7 Sequences of the oligonucleotide primers used to amplify the *HIS3* gene.

C) *Transformation of strain BGY6 with the HIS3 DNA fragment*

Strain BGY6 (Section 3.2.1) was transformed by electroporation (Section 2.2.10d) with approximately 0.2 µg of the (above) purified *HIS3* DNA fragment. Transformed cells were inoculated onto SC-His plates and incubated at 30°C for 3 days.

Overview: The next section describes the conversion of the experimental strain under construction to a Gal^+ phenotype by mating with an appropriate strain, followed by sporulation and spore analysis.



3.3.11 Conversion of the experimental strain to a Gal^+ phenotype

JRY50 is Gal^-; the cause of this phenotype is unknown (J. Rine personal communication). The expression strategy for the I-*Sce*I endonuclease is reliant upon galactose utilisation, the promoter of choice being *GAL1*. It is therefore necessary to convert JRY50 to a Gal^+ phenotype. The mutation causing the Gal^- phenotype could be in any one (or more) genes in the galactose utilisation pathway. Because of this it was decided that the best way to convert JRY50 from a Gal^- to a Gal^+ phenotype would be to mate JRY50 (or more precisely the experimental derivative BGY7) with a suitable Gal^+ strain, followed by sporulation to produce a new haploid strain (BGY8) with the desired characteristics.

The new strain (hereafter BGY8) must of course contain the odorant receptor genes integrated at the donor and acceptor loci, as well as being $HIS3^+$ and GAL^+. It is also important that BGY8 be *MAT***a**, due to the 'recombination proficient' state in which the *HML* locus, i.e. the donor locus, is maintained (Section 1.1.1). BGY8 must also possess one or more auxotrophic markers (e.g. *trp1-289 ura3-52*) to allow selection and maintenance of transformed plasmids (see below). The strain to be crossed, BGY7, also contains the (now) undesired allele *sir4-9*; however this characteristic cannot be easily selected against (see Results).

The steps undertaken towards creating the new Gal^+ strain, BGY8, are described below.

A) *Mating of BGY7 with a Gal^+ strain*

The Gal^+ **α**-strain INVSC2 (Section 3.2.1) was chosen as the strain with which to cross BGY7. This strain contains the same non-reverting *URA3* mutation as BGY7, i.e. *ura3-52*, also suitable markers for the selection of diploid cells following mating, *LEU2 his3Δ200*, which complement/are complemented by (respectively) the auxotrophic markers in BGY6, *leu2Δ::OR17-40 HIS3*, thus allowing diploids to be selected on synthetic medium lacking leucine and histidine.

Fresh BGY7 and INVSC2 cultures were grown on YPD medium (Section 2.2.7) at 30°C. From single colonies of each strain a small amount was transferred - using a sterile cocktail stick - to a fresh YPD plate, where the cells were mixed and spread evenly over a 1 cm^2 area. As a negative control for the mating process both strains were also transferred and spread out individually (no mixing) in separate 1 cm^2 areas on the same YPD plate. The cells were incubated overnight at 30°C.

In order to confirm successful mating and formation of diploids a small amount of incubated cells from each 1 cm^2 patch was transferred (as previously) onto a SC-Leu-His plate (Section 2.2.7), and incubating overnight at 30°C. (Adapted from Guthrie and Fink, 1991).

B) *Sporulation*

Fresh single colonies of the diploid strain (from above) were grown on SC-Leu-His medium (Section 2.2.7). From a single colony, a small number of cells were suspended in water (100 µl), and inoculated onto sporulation medium (Section 2.2.7c). The cells were incubated at 30°C for 4 days to promote spore formation.

C) *Digestion of ascospores*

Spores (from above) were scraped from the sporulation medium using a sterile cocktail stick and resuspended in 50 mM Tris-HCl (200 µl). Lyticase (13 mg/ml in 50 mM Tris-HCl) was added to a final concentration of 2 mg/ml, and the spores incubated at 30°C for 30 minutes, with shaking at 100 r.p.m. The spores were then cooled on ice for 5 minutes before being sonicated in a water bath for 2 minutes. The

spores were then cooled on ice for a further 5 minutes. (Adapted from Campbell and Duffus, 1988.)

D) *Selective killing of vegetative cells*

Diethyl ether (200 µl) was added to the spores (from above) and the suspension shaken by slow vortexing for 10 minutes. The spores were then sat on ice for 5 minutes to allow the layers to separate. The lower aqueous layer was transferred to a fresh polypropylene tube (1.5 ml) and YPD (1.2 ml) added. The spore density of the suspension was determined using a haemocytometer (Section 2.2.11). After making an appropriate dilution, YPD plates (Section 2.2.7) were inoculated with spores to a density of ~500 spores/plate. The spores were incubated for 4 - 5 days at 30°C. (Adapted from Dawes and Hardie, 1974.)

E) *Random spore analysis*

Resultant colonies (from above) were picked - using sterile cocktail sticks - onto gridded YPD plates (Section 2.2.7) in preparation for genetic analysis, and grown as usual. The sizes of the colonies being picked were recorded (as either 'big' or 'small') as an aid in the assignment of diploidy (where big colonies may represent unwanted diploid survivors).

Genetic analysis was performed by transferring each of the colonies growing on the gridded YPD plate onto gridded selective media plates, those being: SC-Leu, SC-Trp, SC-His (Section 2.2.7), in order to select for the auxotrophic markers of interest; and YPG (10g/l Bacto-yeast extract, 20 g/l Bacto-peptone, 20 g/l galactose [added after autoclaving as 20% w/v filter sterilised solution], pH 5 - 6 with 3M NaOH) in order to select for galactose utilisation. The selective media plates were incubated for 3 days at 30°C and the picked colonies inspected for growth.

The colonies were also transferred onto 'mating-type tester' plates in order that mating-type of the colonies could be determined. The 'mating-type tester' plates were YPD plates embedded with either the yeast strain MR9 or RC757 (Section 3.2.1), to a final concentration of 10^4 cells/ml [cultures (10 ml) of both strains were grown overnight and added to the YPD plate medium after it had been allowed to cool to 40°C]. Both 'mating-type tester' strains, MR9 and RC757, are defective in recovery from exposure to the mating pheromones (Section 1.1.1): the **α**-strain RC750 is

unable to grow in the presence of **a**-factor, whilst the **a**-strain, MR9, is unable to grow in the presence of **α**-factor [these *S*uper *S*ensi*T*ive phenotypes are a result of mutations in the genes *SST1* (Dietzel and Kurjan, 1987) and *SST2* (also called *BAR1*; MacKay *et al.*, 1988)]. Thus, the mating-type of picked colonies transferred onto the 'mating-type tester' plates can be identified by a lack of growth (or 'halo') in the underlying embedded cells (following incubation). Diploid cells (that do not produce mating pheromones) are identified by the lack of any 'halo' on either tester plate. The 'mating-type tester' plates were incubated for 2 - 3 days and then inspected for 'halo' formation.

Any colony showing the desired growth characteristics (Section 3.4.10) was further screened by PCR for the presence of the *OR17-228* ORF at the *HML* locus. Screening was performed as described previously (Section 3.3.9b). Glycerol stocks (Section 2.2.8) were prepared from any positive colonies.

3.3.12 The galactose inducible nature of strain BGY8

The efficiency of galactose induction in strain BGY8 (Section 3.2.1) was investigated by assaying β-galactosidase activity (following expression from the galactose inducible vector pYX243). In order that a comparison could be made the Gal$^-$ parent of BGY8, namely BGY7, and the known Gal$^+$ strain FY10 (Section 5.2.1) were similarly assayed.

The three strains were transformed with the pYX243-*LacZ* expression vector (Section 3.2.2g), and assayed for β-galactosidase activity, as described below.

A) *Transformation of BGY7, BGY8 and FY10 with pYX243*

The vector pYX243 [~100 ng] was transformed by electroporation (Section 2.2.10d) into yeast strains BGY7, BGY8 (Section 3.2.1) and FY10 (Section 5.2.1). Leu$^+$ transformants (for each strain) were selected, from which glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). The presence of the vector was confirmed by PCR screening (below).

B) *PCR screening of BGY7, BGY8 and FY10 for pYX243*

Using the primers LacZ(XhB)/Lac(HF) [Table 3.8], and the genomic DNA prepared from each Leu$^+$ strain (above), three *Taq* PCRs (Section 2.2.3a) were performed. The

cycling parameters in each case were: [94°C, 2 min; (94°C, 30s; 50°C, 30 s; 72°C, 30s) × 30].

The bands produced were analysed by agarose gel electrophoresis (Section 2.2.4).

Primer name	Sequence (5' - 3')	Description (5' - 3')
LacZ(XhB)	CGCTAG**CTCGAG**TTAACCAG GCAAAGCGCCATT	5 n.t. spacer; a *Xho*I site; binds to the coding strand of the *LacZ* ORF, 200 - 220 b.p. upstream from the initiation codon.
Lac(HF)	ACCC**AAGCTT**ATGATCCTAGG GCCCCCGT	4 n.t. spacer; a *Hin*dIII site; binds to the template strand of the *LacZ* ORF starting at the initiation codon.

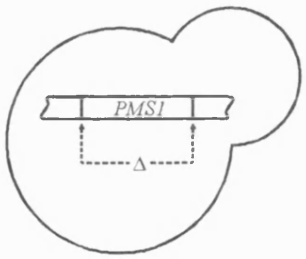
Table 3.8 Sequences of the oligonucleotides used to screen for the presence of the *LacZ* expression vector pYX243. Restriction endonuclease sites are in bold.

C) *Assay for β-galactosidase activity*

A single colony of the strain to be assayed was picked from an SC-Leu plate (2% glucose) into 1 ml of the same medium and grown for 24 hours at 30°C, with shaking at 800 r.p.m. The cells were washed twice with 1 ml of water and resuspended in 1 ml of water. β-galactosidase production was then induced by inoculating 10 ml of fresh liquid SC-Leu (2% galactose) medium with 5×10^7 resuspended cells and incubating for 24 hours at 30°C, with shaking at 250 r.p.m. As a control experiment 5×10^7 resuspended cells were transferred to fresh liquid SC-Leu (2% glucose) medium instead of SC-Leu (2% galactose) medium, and grown as previously. The cell densities of the cultures were determined using a haemocytometer (Section 2.2.11). Approximately $1 - 2 \times 10^7$ cells from each culture were harvested by centrifugation (5 k.r.p.m. for 5 minutes) and resuspended in water (1 ml). The A_{600} for each culture was then determined. The cells were collected by centrifugation (5 k.r.p.m. for 5 minutes) and resuspended in 1 ml of Z-buffer ($Na_2HPO_4.7H_2O$ 90 mM, $NaH_2PO_4.2H_2O$ 35 mM, KCl 10 mM, $MgSO_4.7H_2O$ 1 mM,

β-mercaptoethanol 38 mM). Three drops of chloroform and two drops of 0.1 % SDS were added, and the cells vortexed vigorously for 10 seconds. The sample was then pre-incubated at 28°C for 5 minutes. The assay was started by adding 200 µl of o-nitrophenyl-β-D-galactopyranoside (ONPG; 13 mM in Z-buffer). The reaction was incubated for 1 minute 30 seconds at 28°C, before being stopped by the addition of 1M Na_2CO_3 (0.5 ml). The reaction was centrifuged (13 k.r.p.m. for 1 minute) and the A_{420} of the supernatant measured. β-galactosidase activity was then determined using the formula: $A_{420}/(A_{600} \times$ volume assayed $\times$ time in minutes). [Adapted from Adams *et al.*, 1996.]

Overview: The next section describes the deletion of the *PMS1* gene from the experimental strain under construction.



3.3.13 Creation of a *PMS1* deletion strain

The reason for the creation of an experimental strain in which the *PMS1* locus has been deleted is described in Section 1.3. The same 'pop-in/pop-out' strategy as described previously (Section 3.3.3) will be used again here. The integrating vector once constructed will contain the promoter and terminator regions of the *PMS1* gene, where the intervening *PMS1* ORF has been almost completely deleted (leaving only the last 346 n.t.). Integration of the *PMS1* deletion vector, followed by subsequent excision, should result in a *PMS1*Δ strain.

A) *Primers to amplify the PMS1 promoter and terminator regions*

The primers were designed with reference to the *PMS1* locus sequence, coordinates 473295 - 476009 on Chromosome XIV; obtained from the *Saccharomyces cerevisiae* Genome Database (Section 2.2.1). The primers (Table 3.9) were designed to amplify two approximately 400 b.p. sequences corresponding to the promoter and terminator

regions of the *PMS1* locus. The primers PMS1PF(Xho) and PMS1TB(Bam) incorporated *Xho*I and *Bam*HI restriction sites, respectively, to allow the fragments, once amplified, to be cloned into the integrating vector pRS306 (Section 3.2.2d). The primer PMS1(PB) incorporated homology to the end of the promoter fragment and to the start of the terminator fragment, allowing both fragments, once amplified, to be joined together via a 'splicing by overlap extension' PCR.

Primer name	**Sequence (5' - 3')**	**Description (5' - 3')**
PMS1(PF)Xho	CAGAT**CTCGAG**GCTAGAAAGG GAAGGTTATC	5 n.t. overhang; a *Xho*I site; 20 n.t. homology to the coding strand of the *PMS1* locus, 409 - 389 b.p. upstream from the initiation codon.
PMS1(PB)	CTCTGGAGCCAAACTCTTCTTC TTAGAGAGACGCGTCTTTTC	24 n.t. homology to the template strand of the *PMS1* locus, 346 - 370 b.p. upstream from the termination codon; 20 n.t. homology to the *PMS1* locus, 40 - 20 b.p. upstream from the initiation codon.
PMS1(PF)	AAGAAGAAGAGTTTGGCTCC	20 n.t. homology to the coding strand of the *PMS1* locus, 370 - 350 b.p. upstream from the termination codon.
PMS1(TB)Bam	CAGAT**GGATCC**TGTGTTGTCAC TCCCTGTAT	5 n.t. overhang; a *Bam*HI site; 20 n.t. homology to the template strand of the *PMS1* locus, 65 - 45 b.p. downstream from the termination codon.

Table 3.9 Sequences of the oligonucleotides used in the construction of the *PMS1* knock-out vector pRS306[*PMS1*Δ]. Restriction endonuclease sites are in bold.

B) *PCR amplification of the PMS1 promoter and terminator regions*

Using genomic DNA (Section 2.2.13) from strain BGY8 (Section 3.2.1) two Expand™ PCRs were performed. The first using the primer pair PMS1(PF)Xho/PMS1(PB) to amplify the *PMS1* promoter fragment, and the second using the primer pair PMS1(TF)/PMS1(TB)Bam to amplify the terminator fragment. For both reactions the cycling parameters were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 45 s + 20 s/cycle) × 15; 72°C, 5 min].

Both fragments were analysed by agarose gel electrophoresis (Section 2.2.4), and then purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c).

C) *Splicing of the PMS1 promoter fragment to the PMS1 terminator fragment*

Equal amounts of both fragments (~10 ng), were mixed in an Expand™ PCR with the primers PMS1(PF)Xho and PMS1(TB)Bam. The cycling parameters for the reaction were: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

The spliced PCR product was purified by LMP-agarose gel electrophoresis (Section 2.2.4a) and analysed (Section 2.2.4).

D) *Cloning of the PMS1 spliced product into the integrating vector pRS306*

In separate reactions the spliced *PMS1* fragment (~0.6 µg) and the vector pRS306 (~0.8 µg) were double-digested (Section 2.2.14) with *Bam*HI (15 units) and *Xho*I (10 units) at 37°C overnight. The digested DNAs were purified by LMP-agarose gel electrophoresis (Section 2.2.4a) and recovered by β-agarase digestion (Section 2.2.4b). The purified double-digested products were quantified by agarose gel electrophoresis (Section 2.2.4). The *PMS1* fragment (~40 ng) was then ligated (Section 2.2.15) into the vector (~20 ng), and transformed directly by electroporation into competent *E. coli* (Section 2.2.9d).

Ten ampicillin resistant colonies were PCR screened (Section 2.2.12a) for the presence of the cloned fragment using primers T7 and M13 (Table 5.2). The cycling parameters for the reaction were: [94°C, 5 min; (94°C, 30 s; 48°C, 30s; 72°C, 1 min) × 30]. Positive transformants were grown overnight in 4 ml LB (Section 2.2.6) from

which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

E) *Integration of vector pRS306[PMS1Δ] into yeast strain BGY8*

Prior to transformation the pRS306[*PMS1*Δ] vector (0.5 µg) was linearized by digestion with the restriction enzyme *Eco*RI (40 units), at 37°C overnight (for the same reason as described in Section 3.3.9). The restriction enzyme makes a single cut in the vector within the promoter segment of the cloned *PMS1* fragment, producing free 5'- and 3'-ends that have 225 b.p. and 164 b.p. homology to the target locus respectively. Successful digestion was checked by agarose gel electrophoresis (Section 2.2.4). The cut vector was purified using the Wizard™ DNA Clean-Up System (Section 2.2.4c).

The linearized vector (~0.3 µg) was transformed by electroporation (Section 2.2.10d) into strain BGY8 (Section 3.2.1). The transformed cells were inoculated onto SC-Ura plates (Section 2.2.7a) and incubated at 30°C for 3 days.

Primer name	Sequence (5' - 3')	Description (5' - 3')
PMS1/SCR(B)	ATATTCTTAATGGCTAAGCT	Binds to the coding strand of the *PMS1* locus, 137 - 117 b.p. downstream from the termination codon.
PMS1/SCR(F)	GTGGAACTGTAGGAGCTTCT	Binds to the template strand of the *PMS1* locus, 530 - 510 b.p. upstream from the initiation codon.

Table 3.10 Sequences of the oligonucleotides used to screen for correct integration of vector pRS306[*PMS1*Δ].

F) *Verification of correct integration of plasmid pRS306[PMS1Δ]*

Six Ura^+ colonies were selected from which glycerol stock (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Correct integration of the vector at the *PMS1* locus was first checked by means of a 'straight-through' PCR screen using the

primers PMS1/SCR(F) and PMS1/SCR(B) [Table 3.10]. An Expand™ PCR (Section 2.2.3b) was performed, the cycling parameters for the reaction were: [94°C, 2 min; (94°C, 15 s; 46°C, 30 s; 68°C, 6 min 30 s) × 10; (94°C, 15 s; 46°C, 30 s; 68°C, 6 min 30 s + 20 s/cycle) × 15; 68°C, 10 min].

Both joints (Figure 3.14) formed upon integration were also verified via *Taq* PCRs (Section 2.2.3a). Correct integration of the LHS of the vector into the genome was checked using the primer pair PMS1/SCR(F) [Table 3.10]/T7 [Table 3.2], the cycling parameters were: [94°C, 2 min; (94°C, 30 s; 46°C, 30 s; 72°C, 2 min 30 s) × 30]. Correct integration of the RHS was checked using the primer pair PMS1/SCR(B) [Table 3.10]/M13 (Table 5.2), the cycling parameters were: [94°C, 2 min; (94°C, 30 s; 46°C, 30 s; 72°C, 3 min 30 s) × 30].

The PCR screening products were analysed by agarose gel electrophoresis (Section 2.2.4).

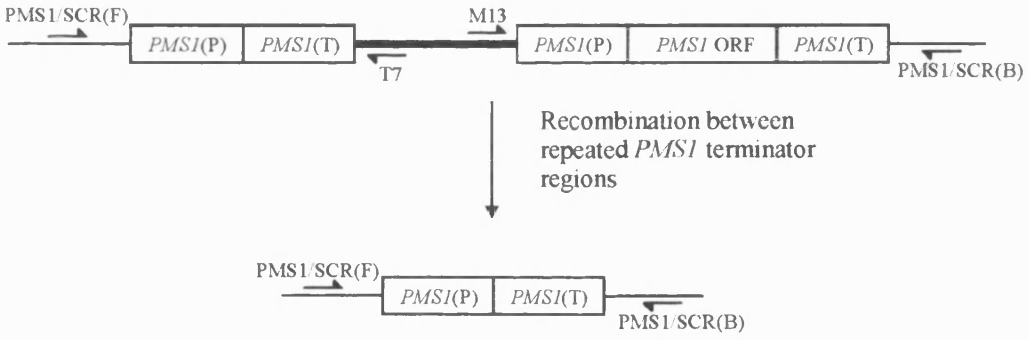


Figure 3.14 Position and orientation of the primers used to verify correct integration (and subsequent loss) of plasmid pRS306[*PMS1*Δ] at the *PMS1* locus. Chromosomal DNA (thin line), pRS306 DNA: thick line. Note that excision of the integrated DNA can proceed via two routes: recombination between the repeated *PMS1* promoter regions, resulting in the wild-type *PMS1* locus (not shown), and recombination between the repeated *PMS1* terminator regions (as shown) resulting in deletion of the *PMS1* ORF (Figure is not to scale).

G) *Selection and screening for vector recombination: loss of URA3 marker*

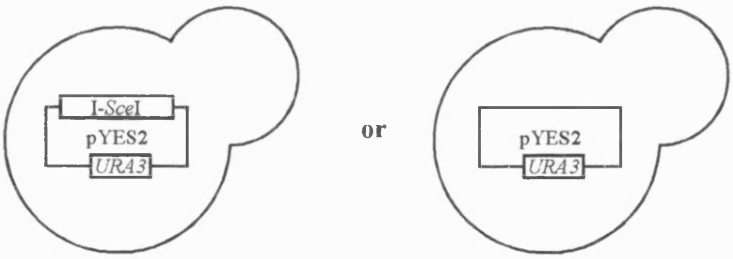
A single positive colony (from the above transformation) was grown overnight at 30°C in SC-Ura (10ml). A loopfull of cells was streaked onto 5-FOA selective medium (Section 2.2.7b) and incubated for 5 days at 30°C. After this period ten large distinct colonies were picked and restreaked (in order to avoid contamination from micro-colony background) onto fresh 5-FOA medium, and allowed to grow for

2 days. Individual colonies were picked into YPD (10 ml) and grown overnight at 30°C. Glycerol stocks (Section 2.2.8) and genomic DNA (Section 2.2.13) were prepared from the resulting cultures.

Genomic DNA prepared from the Ura$^-$ colonies was PCR screened for loss of the pRS306[*PMS1*Δ] vector, resulting in a *PMS1*Δ genotype. A 'straight through' *Taq* PCR (Section 2.2.3a) was performed using the screening primers PMS1/SCR(F) and PMS1/SCR(B). The cycling parameters for the reaction were: [94°C, 2 min; (94°C, 30 s; 46°C, 30 s; 72°C, 3 min) × 30].

PCR products were analysed by agarose gel electrophoresis (Section 2.2.4).

Overview: The next section describes the construction of a vector capable of expressing the I-*Sce*I endonuclease, and its subsequent transformation into the experimental strains under construction. Both experimental strains (the *PMS1* strain and the *pms1*Δ derivative) are also transformed with an expression vector lacking the I-*Sce*I ORF, to create the 'vector-only' controls.



3.3.14 Construction of an I-*Sce*I expression vector

The role of the rare cutter I-*Sce*I endonuclease within the experimental system is described in Section 1.3. The I-*Sce*I expression vector pPEX7 (Section 3.2.2e), containing the universal code equivalent of the I-*Sce*I gene, was supplied by B. Dujon. This vector was unsuitable for use in the experimental strain as the auxotrophic marker present, *LEU2d*, might provide an alternative pathway for repair following double strand cleavage by I-*Sce*I at the chromosomal *LEU2* locus. For this reason the I-*Sce*I ORF was sub-cloned into the expression vector pYES2 (Section 3.2.2b), which contains the alternative marker, *URA3*.

A) *Primers to amplify the I-SceI ORF from vector pPEX7*

The sequence of the I-*Sce*I ORF was obtained from the *Saccharomyces cerevisiae* Genome Database (Section 2.2.1), coordinates 61022 - 61729 (mitochondria). This sequence was then altered to the universal code equivalent (Colleaux *et al.*, 1986). The primers (Table 3.11) to amplify the I-*Sce*I ORF were designed with reference to the universal sequence (Note that there are no differences between the sequences in the regions bound by the primers; see Results). The primers were designed with either an *Eco*RI or a *Xho*I restriction site at their 5'-ends to allow the I-*Sce*I ORF, once amplified, to be cloned into vector pYES2.

B) *Amplification of the I-SceI ORF from vector pPEX7*

Using the primers SceIEcoRI(F)/SceIXhoI(B) [Table 3.11], and the pPEX7 vector (Section 3.2.2f) as template, an Expand™ PCR was performed with the following cycling parameters: [94°C, 2 min; (94°C, 15 s; 38°C, 1 min; 72°C, 1 min) × 10; (94°C, 30 s; 38°C, 1 min; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

The amplified fragment was analysed by agarose gel electrophoresis (Section 2.2.4), and purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c).

C) *Sub-cloning of the I-SceI ORF into pYES2*

The (above) amplified I-*Sce*I ORF fragment (~0.3 μg), and the vector pYES2 (~10 ng), were separately double-digested with the enzymes *Xho*I (30 units) and *Eco*RI (30 units), overnight at 37°C. The digested DNAs were purified by LMP-agarose gel electrophoresis (Section 2.2.4a), and recovered from the gel by β-agarase digestion (Section 2.2.4b). The purified double-digested products were quantified by agarose gel electrophoresis (Section 2.2.4). The I-*Sce*I ORF fragment (~30 ng) was then ligated (Section 2.2.15) into the vector (~10 ng), and transformed directly by electroporation into competent *E. coli* (Section 2.2.9d).

Ten ampicillin resistant colonies were PCR screened (Section 2.2.12a) for the presence of the cloned fragment using primers Shuffle(F) and Shuffle(B) [Table 3.11]. The cycling parameters for the reaction were: [94°C, 5 min; (94°C, 30 s; 48°C, 30s; 72°C, 1 min) × 30]. Positive transformants were grown overnight in 4 ml LB

(Section 2.2.6) from which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

Primer name	Sequence (5' - 3')	Description (5' - 3')
SceIEcoRI(F)	CAGAT**GAATTC**ATGAAAA ATATTAAAAAAAAATCAA	5 n.t. overhang; an *Eco*RI site; 25 n.t. homology to the start of the I-*Sce*I ORF (coding strand).
SceIXhoI(B)	GGCTAG**CTCGAG**TTATTTT AAAAAAGTTTCGG	6 n.t. overhang; a *Xho*I site; 20 n.t. homology to the end of the I-*Sce*I ORF (template strand).
Shuffle(F)	TACTTTAACGTCAAGGAG	Binds to the pYES2 vector, 76 - 58 b.p. downstream from the end of the multiple cloning site.
Shuffle(B)	TTCGGTTAGAGCGGATGT	Binds to the pYES2 vector, 70 - 52 b.p. upstream from the start of the multiple cloning site.

Table 3.11 Sequences of the oligonucleotides used to amplify the I-*Sce*I ORF from vector pPEX7, and those used to confirm successful sub-cloning of the fragment. Restriction endonuclease sites are in bold.

D) *Verification of the sub-cloned I-SceI ORF by DNA sequencing*

Plasmid DNA from one positive colony (above) was sequenced (Section 2.2.16). Two sequencing reactions were performed using the primers Shuffle(F) and Shuffle(B) [Table 3.11].

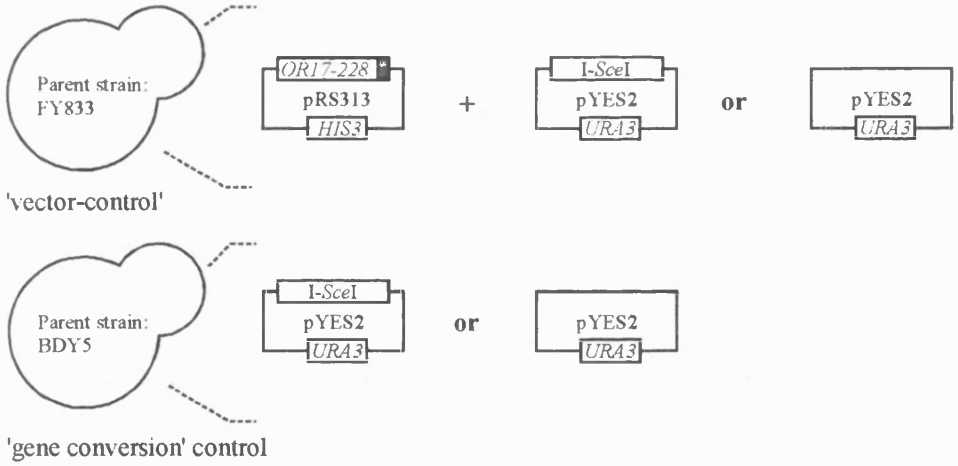
3.3.15 Transformation of the I-*Sce*I expression vector pYES2[I-*Sce*I] into the experimental strains BGY8 and BGY10

The (above) vector pYES2[I-*Sce*I] (~0.1 µg) was transformed by electroporation (Section 2.2.10d) into the experimental strains BGY8 and BGY10 (Section 3.2.1). In the same manner both strains were also transformed with the vector pYES2 only

(Section 3.2.2b) [for use as negative control strains where I-*Sce*I endonuclease expression cannot occur]. Transformants were inoculated onto SC-Ura plates (Section 2.2.7) and incubated for 3 days at 30°C.

Ura^+ transformants were PCR screened (Section 2.2.12b) for the presence of the vector using the primers Shuffle(F) and Shuffle(B) [Table 3.11]. The cycling parameters for the reaction were: [94°C, 5 min; (94°C, 30 s; 48°C, 30s; 72°C, 1 min) × 30]. Glycerol stocks (Section 2.2.8) were made of positive transformants.

Overview: The next section describes the construction of a plasmid containing the I-*Sce*I recognition site and its transformation (along with the I-*Sce*I expression vector or the 'vector-only') into parent strain FY833, to generate the 'vector control' strains. The next section also describes the construction of the 'gene conversion' control strains by the transformation of strain BDY5 with the I-*Sce*I expression vector, or with the 'vector-only'.



3.3.16 Construction of control strains

Two control strains were constructed. The first strain, containing the I-*Sce*I expression vector pYES2[I-*Sec*I] and a second vector bearing the I-*Sce*I recognition site, was a means to establish that I-*Sce*I induced DNA cleavage was occurring as expected. The second strain was to act as a control to the experimental strains (and the gene conversion process), this strain contains a similar acceptor locus arrangement (described below) to the experimental strains, but does not contain a donor locus. This strain also contained the I-*Sce*I expression vector pYES2[I-*Sce*I]. As for the

experimental strains (above) both control strains were also constructed containing the pYES2 only vector.

The construction of both control strains, and of a vector containing the I-*Sce*I restriction site, are described below.

A) *Construction of a vector containing the I-SceI restriction site*

This was achieved by amplifying a segment from a previously made DNA fragment (containing the I-*Sce*I site) and cloning of the fragment into a suitable vector. Thus, the primers Leu2XbaI (Table 5.5) and OR17-228(FX) [Table 3.12] were used to amplify the I-*Sce*I site (and the proceeding *OR17-228* ORF) from the previously made *HML* disruption cassette (Section 3.3.6d). The primers introduced the restriction sites *Xho*I and *Xba*I at either end of the fragment for ease of cloning. The *HML* disruption cassette is shown again, with the positions of the primers, in Figure 3.15. An Expand™ PCR (Section 2.2.3b) was performed with the following cycling parameters: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

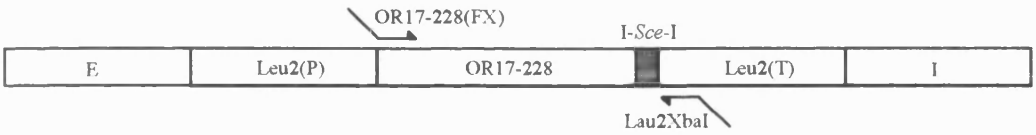


Figure 3.15 The *HML* locus gene disruption cassette showing the positions of the primers used to amplify a fragment containing the I-*Sce*I recognition site for cloning into vector pRS313 (Figure is not to scale).

The amplified product was analysed as usual (Section 2.2.4), then double-digested (Section 2.2.14) with restriction enzymes *Xba*I (10 units) and *Xho*I (10 units), at 37°C overnight. The digested fragment was then purified by LMP-agarose gel electrophoresis (Section 2.2.4a). The vector pRS313 (Section 3.2.2h) was similarly digested and purified. The insert (~50 ng) was ligated (Section 2.2.15) into the cut vector (~20 ng) and transformed directly into *E. coli* (Section 2.2.9d).

Ten ampicillin resistant colonies were PCR screened (Section 2.2.12a) for the presence of the cloned fragment using primers T7 and M13 (Table 5.2). The cycling

parameters for the reaction were: [94°C, 5 min; (94°C, 30 s; 50°C, 30s; 72°C, 1 min) × 30]. Positive transformants were grown overnight in 4 ml LB (Section 2.2.6) from which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

Plasmid DNA from one positive colony was sequenced in order to check the integrity of the cloned I-*Sce*I site. Two sequencing reactions (Section 2.2.16) were performed using the primers M13 and T7 (Table 5.2).

Primer name	Sequence (5' - 3')	Description (5' - 3')
OR17-228(FX)	CGATCC**CTCGAG**ATGGAGCCA GAAGCTGGGAC	6 n.t. overhang; a *Xho*I site; 20 n.t. homology to coding strand of *OR17-228* ORF starting at the initiation codon.

Table 3.12 Sequence and description of the oligonucleotide primer OR17-228(FX). Restriction endonuclease site is shown in bold.

B) *Construction of a strain to show I-SceI induced DNA cleavage*

Two sequential transformation were performed to introduce the vectors of interest into yeast strain FY833 (Section 3.2.1). In the first transformation, the engineered vector (from above) containing the I-*Sce*I restriction site, pRS313[I-*Sce*I], was introduced. Transformants were selected on SC-His plates (30°C for 3 days). His^+ transformants were PCR screened for the presence of the vector as described in Section 3.3.16a, positive colonies were then transformed with the I-*Sce*I expression vector, pYES2[I-*Sce*I]. His^+ colonies were also transformed with the control vector pYES2 only. In both cases transformants were selected on SC-His-Ura plates (30°C for 3 days). His^+ Ura^+ transformants were PCR screened for the presence of the vector pYES2[I-*Sce*I] or pYES2 only, as described in Section 3.3.15. A third transformation was also performed where the His^+ strain (containing the vector pRS313[I-*Sce*I]) was transformed with the original I-*Sce*I expression vector, pPEX7 (Section 3.2.2e). Transformants were selected on SC-His-Leu plates (30°C for 3 days) [no screening was performed, see Results]. For each transformation described, approximately 0.1 µg of DNA was used.

C) *Construction of the gene conversion control strain*

The strain used to construct the gene conversion control strain was BDY5 (Section 5.2.1). This strain was constructed in the course of the work presented in Chapter 5 and is described in detail therein. Briefly, BDY5 is a Gal^+ strain possessing a wild-type *HML* locus. Integrated between two *lox* sites at the *LEU2* locus is the *OR17-228* ORF followed by the I-*Sce*I recognition site (in the same orientation as the experimental strains described in this chapter)].

In two separate transformations, strain BDY5 was transformed with vectors pYES2[I-*Sce*I] and pYES2 only. In both cases approximately 0.1 μg of DNA was used. Transformants were inoculated onto SC-Ura plates (Section 2.2.7) and incubated for 3 days at 30°C.

Ura^+ transformants were PCR screened for the presence of the vectors as described in Section 3.3.15. Glycerol stocks (Section 2.2.8) were made of positive transformants.

3.4 RESULTS: GENERATION OF EXPERIMENTAL STRAINS FOR THE ANALYSIS OF MOLECULAR EVOLUTION BY GENE CONVERSION

3.4.1 Construction of a cassette for the integration of *OR17-40* at the acceptor locus *LEU2*

The main steps undertaken in the construction of the disruption cassette are shown diagrammatically in Figure 3.9.

The first step involved the amplification of two regions of the *LEU2* locus to flank the transgene (the *OR17-40* ORF) in the left and right hand side direct repeat sequences. Each region of the *LEU2* locus was amplified twice (as shown) in order that different 5'- and 3'- ends could be incorporated into each direct repeat (see below). The four fragments of interest were amplified from yeast genomic DNA (Section 3.3.1b), and the sizes of the fragments determined by agarose gel electrophoresis. Each fragment was in accordance with that expected, thus: Leu2 promoter fragment for the LHS repeat sequence, 344 b.p., and for the RHS repeat, 324 b.p.; Leu2 terminator fragment for the LHS repeat sequence, 406 b.p., and for the RHS repeat, 324 b.p. An I-*Sce*I restriction site was incorporated into the *LEU2* terminator fragments by the primer LeuTF.

The *OR17-40* open reading frame was amplified from vector pYES2[*OR17-40*] (Section 3.3.1c). The size of the generated PCR fragment was in accordance with that expected: 948 b.p.

The appropriate *LEU2* fragments were joined to the *OR17-40* open reading frame by means of a 'splicing by overlap extension' reaction. Each repeat sequence, i.e. LHS and RHS, was generated in a separate reaction, the sizes of the products were as expected, 1650 b.p. for the LHS, and 1634 b.p. for the RHS.

The fragments generated, as indicated by their slight differences in size, were not identical. In order that the repeat sequences could be (in the subsequent step) joined to the *K. lactis URA3* gene, the LHS repeat incorporated a *Bgl*I site at its 3'-end (introduced by the primer LeuTB/Bgl), and the RHS repeat a *Bgl*I site at its 5'-end (introduced by the primer LeuPF/Bgl). Also, at the opposite ends to the *Bgl*I sites, each fragment incorporated ~20 n.t. of unique sequence, i.e. the Leu2(P) fragment and the Leu2(T) fragment at the extreme 5'-end and 3'-ends of the completed cassette

contained sequences not present at the 'internal' copies of these fragments. This was to enable the completed disruption cassette to be amplified by PCR, using primers that bind only to the unique sequences at the ends of the fragment.

The *K. lactis URA3* selectable marker was amplified from vector pTG5756 (Section 3.3.1f), producing a fragment (1448 b.p.) with a *Bgl*I site at each end. The completed disruption cassette was generated by ligating the *URA3* fragment (following *Bgl*I digestion) to the LHS and RHS repeat sequences (similarly digested with *Bgl*I, Section 3.3.1g). The completed cassette (4690 b.p.) is shown in lane 2, Figure 3.16.

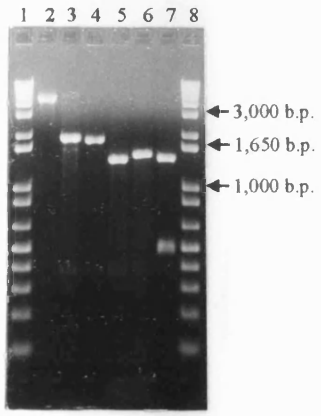


Figure 3.16 1% agarose gel showing the completed *LEU2* gene disruption cassette (Lane 2; 4690 b.p.) and various bands confirming correct cassette integration: Lanes 1 and 8, 1 k.b.p. DNA marker (Figure 2.1); Lane 3, the wild-type *LEU2* locus (1930 b.p.); Lane 4, *OR17-40* ORF at the *LEU2* locus (1847 b.p.); Lane 5, the LHS novel junction at the *LEU2* locus (1348 b.p.); Lane 6, the RHS novel junction at the *LEU2* locus (1447 b.p.); and Lane 7, the bands produced (482 b.p. and 1365 b.p.) following digestion of the DNA fragment (lane 4) with I-*Sce*I endonuclease.

3.4.2 Integration of the *OR17-40* ORF at the *LEU2* locus of *Saccharomyces cerevisiae* strain JRY50

The gene disruption cassette (previous section) was transformed into yeast strain JRY50 using a lithium acetate/polyethyleneglycol transformation procedure (Section 3.3.2 a). After three days incubation ~20 colonies were observed growing on each plate, corresponding to approximately 60 Ura^+ transformants per 0.1 μg of DNA.

A variety of colony sizes were observed, with most of the colonies being either small or medium in size. Six of the largest colonies were selected and screened by PCR in order to confirm correct cassette integration (Section 3.3.2b); the screens were designed to amplify across the two novel junctions created (Figure 3.10). In only one of the six colonies did the screen produce the expected PCR products for both reactions, i.e. 1779 b.p. for the LHS junction, and 1813 b.p. for the RHS junction. This colony, containing the disruption cassette correctly integrated at the *LEU2* locus, was designated BGY1 (Section 3.2.1).

Following overnight growth in non-selective medium, approximately 2.5×10^6 BGY1 cells were plated onto 5-FOA medium, in order to select Ura^- colonies. After five days growth each selection plate (4 in total) contained approximately 500 distinct colonies, giving a recombination frequency of $\sim 80^{-4}$ (close to the 10^{-4} reported by Alani *et al.*, 1987; Section 3.3.1). Eight colonies were PCR screened (Section 3.3.2c) in order to ascertain that recombination had occurred as intended (Figure 3.10). A 'straight through' screen revealed that the 'full' disruption cassette was no longer present in any of the colonies; each producing a band of 1847 b.p. (lane 4, Figure 3.16), that expected if the *OR17-40* ORF only was present at the *LEU2* locus (the wild-type *LEU2* locus [1930 b.p.] is shown for comparison in lane 3, Figure 3.16; the 83 b.p. difference between the two sequences being just discernible). The two 'new' novel junctions created were also screened, once again all of the colonies produced the expected bands: 1348 b.p. for the LHS junction (lane 5, Figure 3.16) and 1447 b.p. for the RHS junction (lane 6, Figure 3.16). The new strain was designated BGY2 (Section 3.2.1).

The integrity of the integrated *OR17-40* ORF was further verified by DNA sequencing (Section 3.3.2d) in order to check for mutations that might have arisen during construction of the disruption cassette. Four mutations were noted in the sequence (Table 3.13).

Mutation	Position (downstream) w.r.t. *OR17-40* initiation codon
T→G	13
G→A	19
C→A	934
G→A	936

Table 3.13 Mutations found in the *OR17-40* ORF integrated at the *LEU2* locus in BGY2.

The mutations at positions 13 and 19 were introduced by the degenerate primer LeuTB (Table 3.1) and those at positions 934 and 936 by the degenerate primer LeuTF (Table 3.1). These primers, LeuTB and LeuTF, were designed so that an amplified sequences (the *LEU2* promoter or the *LEU2* terminator fragment) could be spliced to both the *OR17-40* and *OR17-228* open reading frames, which differ from each other by four bases in the 21 n.t. region incorporated into the LeuTF primer, and by three bases in the 24 n.t. region incorporated into the LeuTB primer.

Sequencing also revealed the I-*Sce*I recognition site to be free from mutations. The viability of the site was further demonstrated by digesting the DNA fragment amplified for sequencing purposes (lane 4, Figure 3.16), with the I-*Sce*I endonuclease (Section 3.3.2d). After 1 hours digestion the two expected bands, 482 b.p. and 1365 b.p. were produced (lane 7, Figure 3.16).

3.4.3 Construction of the *E-URA3-I HML* disruption cassette

The *E* and *I* flanking regions (Figure 3.11) were individually amplified from yeast genomic DNA (Section 3.3.4b), and the sizes of the fragments determined by agarose gel electrophoresis. Both fragments were in accordance with that expected, 428 b.p. for the *E* DNA fragment and 535 b.p. for the *I* DNA fragment. The disruption cassette was generated by joining the *E* and *I* DNA fragments to the previously amplified *K. lactis URA3* gene by means of a 'splicing by overlap extension' reaction. The size of the completed cassette when checked by agarose gel electrophoresis was as expected: 2363 b.p.

3.4.4 Integration of the *E-URA3-I HML* disruption cassette into *Saccharomyces cerevisiae* strain BGY2

The *E-URA3-I* disruption cassette (previous section) was transformed into yeast strain BGY2 by electroporation (Section 3.3.5a). After three days incubation a total of twelve colonies were observed growing on the transformation plates.

All 12 Ura^+ colonies were PCR screened in order to confirm correct cassette integration (Section 3.3.5b). The screens were designed to amplify across the two novel junctions created (Figure 3.12). One colony produced the expected screening bands for both reactions, i.e. 611 b.p. for the LHS junction, and 678 b.p. for the RHS

junction. This colony, containing the *K. lactis URA3* gene integrated at the *HML* locus, was designated BGY3 (Section 3.2.1).

The *E-URA3-I* disruption cassette was also transformed and successfully integrated (with similar results to above) into the parent strain JRY50, producing strain BGY4 (Section 3.2.1). [The growth characteristics of strain BGY4, growing on SC-Ura medium, were investigated at 24°C and 34°C. It was expected that the strain would not be able to grow at the lower temperature due to silencing of the *URA3* gene at *HML*, as described in Section 3.3.3. However, no clear effect was observed. The colonies grew more slowly at 24°C, but no more slowly than JRY50 only (growing on SC medium). This result (and the *sir4-9* effect observed during spore analysis; Section 3.4.10) are considered further in the Discussion (Section 3.5.1).]

3.4.5 Construction of the *E-OR17-228-I* gene replacement cassette

The steps involved in the construction of the gene replacement cassette are shown diagrammatically in Figure 3.11b. In the first step, the construction of the 'central portion', the *OR17-228* ORF - amplified from vector pYES2[*OR17-228*] (Section 3.3.6a) - was joined via a 'splicing by overlap extension' reaction (Section 3.3.6b) to the previously made *LEU2*-promoter and *LEU2*-terminator fragments. The spliced product was of the expected size (1658 b.p.) when analysed by agarose gel electrophoresis.

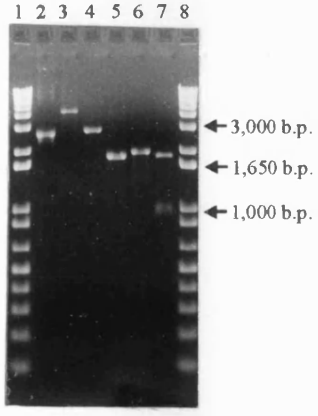


Figure 3.17 1% agarose gel showing the completed gene replacement cassette (Lane 2; 2573 b.p.) and various bands confirming correct cassette integration: Lanes 1 and 8, 1 k.b.p. DNA marker (Figure 2.1); Lane 3, the wild-type *HML* locus (4081 b.p.); Lane 4, the gene replacement cassette integrated at the *HML* locus (2752 b.p.); Lane 5, the LHS novel junction at the *HML* locus (1792 b.p.); Lane 6, the RHS novel junction at the *HML* locus (1914 b.p.); and Lane 7, bands produced (950 b.p. and 1805 b.p.) following digestion of the DNA fragment (lane 4) with I-*Sce*I endonuclease.

The *E* and *I* flanking regions were individually amplified from yeast genomic DNA (Section 3.3.6c). Both fragments were of correct size when analysed, 428 b.p. for the *E* DNA fragment and 535 b.p. for the *I* DNA fragment.

The three DNA fragments were joined to one another via a 'splicing by overlap extension' reaction (Section 3.3.6d). The size of the completed cassette was in accordance with that expected, 2573 b.p. The amplified fragment is shown in lane 2, Figure 3.17.

3.4.6 Integration of the *E-OR17-228-I* cassette into BGY3: direct gene replacement

Four attempts were made to transform the *E-OR17-228-I* cassette directly into strain BGY3; three times by electroporation and once using the lithium acetate/polyethyleneglycol transformation protocol (Section 3.3.7a). Each attempt produced a few colonies (mostly small and slow growing) on the 5-FOA selective medium. However, when screened (Section 3.3.7b), none of the colonies produced the expected sized band corresponding to cassette integration (2757 b.p.), instead producing a band indicative of the continued presence of the *URA3* gene (2547 b.p.). These 5-FOA resistant colonies presumably arose due to spontaneous genomic mutations.

3.4.7 Construction of the integrating vector pRS306[*OR17-228*]

A modified gene replacement cassette was produced, incorporating a *Xho*I restriction site at its 5'-end and a *Xba*I restriction site at its 3'-end. The fragment was double-digested with the *Xho*I and *Xba*I enzymes and ligated into the integrating vector pRS306, bearing compatible cohesive ends. The ligation was transformed directly into competent *E. coli* cells. Twelve ampicillin resistant colonies were PCR screened, of which 2 were found to contain the insert, producing a fragment corresponding with that expected: 2740 b.p. The new vector, pRS306[*OR17-228*], was as for pRS306 (Section 3.2.2d), except that the *OR17-228* gene replacement cassette (Figure 3.11b) was cloned at the multiple cloning site between the *Xho*I and *Xba*I restriction sites.

3.4.8 Transformation of the integrating vector pRS306[*OR17-228*] into BGY2

The linearized vector was transformed by electroporation into strain BGY2 (Section 3.3.9b). After three days growth 17 Ura^+ colonies were obtained, corresponding to a transformation efficiency of approximately 500 colonies/0.1μg DNA.

Six Ura^+ colonies were screened via a 'straight through' PCR (Section 3.3.9a), two colonies produced the ~10 k.b.p. band corresponding to the vector integrated at the *HML* locus. The other four colonies produced a band corresponding to the wild-type *HML* locus (~4 k.b.p.).

The novel joints formed upon vector integration were further verified in one of the colonies. Both screens produced the expected sized products when analysed by agarose gel electrophoresis: 2747 b.p. for the LHS junction, and ~4 k.b.p. for the RHS junction. The new strain, containing pRS306[*OR17-228*] integrated at the *HML* locus, was designated BGY5.

Strain BGY5 was streaked onto 5-FOA medium and incubated at 30°C for 5 days. After this period 8 distinct colonies were observed growing, from which glycerol stocks and genomic DNA were prepared. A 'straight through' PCR screen of the colonies (Section 3.3.9b) revealed that for each one the vector sequence had been excised: in 4 of the colonies vector excision had resulted in reversion to the wild-type locus, producing a band of 4081 b.p. (lane 3, Figure 3.17); whilst in the remaining colonies vector excision has produced the desired integration event, producing a band of 2757 b.p. (lane 4, Figure 3.17), corresponding to the *OR17-228* ORF (flanked by *LEU2* sequences) at the *HML* locus.

In one colony, correct vector excision was further verified by checking both novel joints created: A band of 1792 b.p. (lane 5, Figure 3.17) was produced for the LHS screen and 1914 b.p. (lane 6, Figure 3.17) for the RHS screen; both corresponding to those expected.

The new strain, containing the *OR17-40* ORF at the acceptor locus, and the *OR17-228* ORF at donor locus, was designated BGY6 (Section 3.2.1).

Verifying the integrity of the integrated DNA sequence

The integrity of the *OR17-228* ORF integrated in strain BGY6 was further verified by DNA sequencing (Section 3.3.9c). Five mutations were noted in the sequence (Table 3.14).

Mutation	Position (downstream) w.r.t. *OR17-228* initiation codon
A→G	487
G→A	828
A→G	936
A→G	943
C→T	945

Table 3.14 Mutations found in the *OR17-228* ORF integrated at the *HML* locus in BGY6.

As for the mutations found at the end of the *OR17-40* ORF (Section 3.4.2), the mutations at positions 936, 943, and 945, in the *OR17-228* ORF, are again due to the degenerate primer LeuTF (Table 3.1), for the reason described previously. The A→G and G→A mutations at positions 487 and 828, respectively, represent polymerase errors in replication that have arisen during construction of the integrating vector.

Sequencing showed the I-*Sce*I recognition site after the *OR17-228* ORF to be mutation free. As for the I-*Sce*I site incorporated after the *OR17-40* ORF, the viability of the site after the *OR17-228* ORF was demonstrated by digesting the DNA fragment amplified for sequencing purposes (lane 4, Figure 3.17) with the I-*Sce*I endonuclease (Section 3.3.9c). After 1 hours digestion the two expected bands, 950 b.p. and 1805 b.p. were produced (lane 7, Figure 3.17).

3.4.9 Conversion of the experimental strain to a His^+ phenotype

A *HIS3*-containing DNA fragment was amplified from the yeast strain FY10 (Section 3.3.10b). When analysed by agarose gel electrophoresis the fragment size was in accordance with that expected: 1937 b.p. Following purification, the fragment was transformed directly into strain BGY6 (Section 3.3.10). After 3 days growth approximately 200 His^+ colonies were obtained, corresponding to a transformation efficiency of approximately 600 colonies/0.1 μg of DNA.

Due to the untargeted nature of this transformation, it was not possible to verify integration of the transformed DNA fragment at the *HIS3* locus (see Discussion). However, the transformed cells exhibited the desired phenotype - they were able to grow on medium lacking histidine - and were therefore selected for use in the next step in the construction of the experimental strain: conversion to a Gal^+ phenotype. The new His^+ strain was designated BGY7 (Section 3.2.1).

3.4.10 Conversion of the experimental strain to a Gal^+ phenotype

For the reason described in Section 3.3.11 the experimental strain BGY6 was crossed with the Gal^+ strain INVSC2 (Section 3.3.11a). Successful mating was demonstrated by the ability of the resultant diploid cells to grow on SC-Leu-His selective medium. Neither BGY6 nor INVSC2 alone were able to grow on this medium.

The diploid cells were sporulated by incubating on 1.5% potassium acetate medium (Section 3.3.11b). After 4 days incubation approximately 50% spore formation was observed. Ascospores were scraped from the sporulation plate and subjected to digestion with lyticase and sonication (in order to release individual spores) [Section 3.3.11c]. The spores, having been released from their ascus, were then subjected to diethyl ether treatment in order to kill any vegetative cells (Section 3.3.11d). The spores were plated onto YPD medium and incubated. Approximately 5% of the inoculated spores grew to give colonies. [The low number of spores growing to give colonies was not a problem. Dawes and Hardie (1974) reported 50% spore survival following optimisation of exposure time to diethyl ether.]

Eighty-five colonies were picked for analysis onto the various selective media (Section 3.3.11e). Of these, 29 did not show any 'halo' on either mating-type tester plate, indicating that they were diploid. These colonies were discarded from the segregation results presented below. No correlation was observed between diploidy and those colonies initially characterised as being either 'big' or 'small'. The diploid colonies (above) most probably do not represent surviving (pre-sporulation) vegetative cells; instead they are most likely to have arisen by mating of individual spores (some remain associated after treatment with lytic enzyme and sonication). This is evident in the fact that only one of the diploid colonies was Gal^+ His^+ Leu^+ and Trp^+. The segregation analysis of the remaining 56 haploid colonies is shown in Table 3.15.

Genotype (a-mating type colonies)	**Number of colonies (24°C)**	**Number of colonies (34°C)**	**Genotype (α-mating type colonies)**	**Number of colonies (24°C)**	**Number of colonies (34°C)**
$Leu^+ Trp^+ His^+ Gal^+$	2	2	$Leu^+ Trp^- His^+ Gal^+$	2	1
$Leu^+ Trp^+ His^- Gal^+$	2	2	$Leu^+ Trp^- His^- Gal^-$	2	2
$Leu^+ Trp^+ His^- Gal^-$	2	2	$Leu^- Trp^- His^+ Gal^+$	1	0
$Leu^- Trp^+ His^- Gal^-$	1	1	$Leu^+ Trp^+ His^- Gal^+$	2	1
$Leu^- Trp^+ His^+ Gal^-$	4	3	$Leu^- Trp^- His^+ Gal^-$	2	1
$Leu^- Trp^+ His^+ Gal^+$	2*	2	$Leu^+ Trp^+ His^+ Gal^+$	4	3
$Leu^- Trp^- His^- Gal^+$	1	0	$Leu^+ Trp^+ His^- Gal^-$	2	0
$Leu^+ Trp^- His^- Gal^-$	3	3	$Leu^- Trp^- His^- Gal^-$	1	1
$Leu^+ Trp^+ His^+ Gal^-$	3	3	$Leu^+ Trp^- His^+ Gal^-$	3	3
$Leu^+ Trp^- His^- Gal^+$	1	1	$Leu^- Trp^+ His^- Gal^-$	2	1
$Leu^+ Trp^- His^+ Gal^-$	4	4	$Leu^- Trp^+ His^- Gal^+$	1	1
$Leu^- Trp^- His^+ Gal^+$	3*	3	$Leu^+ Trp^+ His^+ Gal^-$	1	1
$Leu^- Trp^- His^+ Gal^-$	2	2		**23 total**	**15 total**
$Leu^- Trp^- His^- Gal^-$	3	3			
	33 total	**31 total**			

Table 3.15 Segregation analysis results of 56 haploid colonies. Colonies are divided into two main groups according to mating-type (**a** or **α**). Those colonies marked with an asterisk (*) have the desired characteristics (as described in Section 3.3.11), those being: **a**-mating type, Leu^- (due to presence of *OR17-40* at locus), His^+, Gal^+ and either Trp^- or Trp^+. Analysis of mating-type was performed at two growth temperatures - 24 and 34°C - in order to look for the effect of the *sir4-9* mutation. The temperature sensitive *sir4-9* mutation is manifested as a 'pseudo-diploid' phenotype at 34°C (see main text), resulting in fewer haploid colonies being recorded.

Five of the colonies had the desired combination of attributes (before PCR screening). However, when cultures of these colonies were prepared - with the aim of preparing genomic DNA from which to screen the *HML* locus - a previously unknown (and unwanted) phenotype was observed. The cells in all but one of the cultures exhibited flocculating growth (cells growing in clusters). This was found (by microscope inspection) to be a property of the parent strain, INVSC2.

The *HML* locus of the one colony showing normal growth was PCR screened for the desired sequences, all of which were shown to be present and correct. The new Gal^+ strain (also Trp^+) was designated BGY8 (Section 3.2.1).

The earlier growth study using strain BGY4 (Section 3.4.4) had shown no discernible temperature dependent growth pattern due to the *sir4-9* mutation. However, in the analysis of the mating-type of the (above) colonies, a clear temperature dependent effect was observed. At 34°C the number of diploid colonies counted was more than when the same colonies were counted at 24°C (Table 3.15). This effect, observed in 8 of the 23 *MAT*α colonies, and in 2 of the 33 *MAT***a** colonies, was as expected (due to colonies containing the *sir4-9* mutation). At 34°C the silenced loci in these colonies are derepressed, resulting in a 'pseudo-diploid' phenotype, due to expression of both sets of mating-type genes. This effect is reversed at the lower temperature, when the *HML* and *HMR* loci are again silenced. The greater number of 'pseudo-diploids' amongst *MAT*α colonies compared to *MAT***a** colonies was also as expected. There should be twice as many 'pseudo-diploid' *MAT*α colonies as *MAT***a** colonies, as two possible *MAT*α strains: *HML*α *MAT*α *HMR***a** *sir4-9* and *hml*Δ::*ORI7-228 MAT*α *HMR***a** *sir4-9*; but only one possible *MAT***a** strain: *HML*α *MAT***a** *HMR***a** *sir4-9*; can exhibit the 'pseudo-diploid' phenotype. The second *MAT***a** strain, *hml*Δ::*ORI7-228 MAT***a** *HMR***a** *sir4-9* (which may correspond to BGY8), does not show this effect, as derepression of *HML* and *HMR* still results in a *MAT***a** phenotype, hence the nature of the *SIR4* allele in BGY8 remains unknown (this problem, and the above effect observed due to the *sir4-9* allele, are considered further in the Discussion).

3.4.11 The galactose inducible nature of strain BGY8

The *LacZ* control vector pYX243 was transformed by electroporation into strain BGY8, and also into the Gal^- parent strain BGY7 and the known Gal^+ strain FY10 (Section 3.3.12a). After three days incubation numerous Leu^+ colonies were observed for each strain. A single colony of each strain was PCR screened for the presence of the vector (Section 3.3.12b), each colony produced the expected 210 b.p. screening band. The three new strains were designated BGY8Z, BGY7Z and FY10Z (Section 3.2.1).

The β-galactosidase assay was performed on each strain as described in Section 3.3.12c. The results from two experiments are presented in Table 3.16.

Strain	BGY7Z	BGY8Z	FY10Z
Activity	1. 0.19	1. 0.20	1. 0.25
$(gmol^{-1})_{A420}(gmol^{-1})^{-1}{}_{A600}\ ml^{-1}\ min^{-1}$	2. 0.20	2. 0.16	2. 0.27

Table 3.16 β-galactosidase activity for galactose induced cultures. For each strain (and both experiments) no activity was observed for the glucose controls.

The validity and meaning of these results are discussed in Section 3.5.1.

3.4.12 Creation of a *PMS1* deletion strain

Two fragments, corresponding to the promoter and terminator regions of the *PMS1* gene, were individually amplified from genomic DNA (Section 3.3.13b). Both fragments were as expected when analysed: 432 b.p. for the promoter fragment and 424 b.p. for the terminator fragment. The fragments were jointed to one another by means of a 'splicing by overlap extension' PCR (Section 3.3.13c). The spliced product when analysed was in accordance with that expected: 832 b.p.

The *PMS1* promoter/terminator fragment was double-digested with *Xho*I and *Bam*HI enzymes and ligated into the similarly digested pRS306 vector (Section 3.3.13d). The ligation was transformed directly into competent *E. coli* cells. Ten ampicillin resistant colonies were PCR screened, 5 of which produced the expected screening band of 952 b.p., corresponding to correctly cloned insert. The new vector, pRS306[*PMS1*Δ], was as for pRS306 (Section 3.2.2d), except that the *PMS1* promoter/terminator fragment (Section 3.3.13d) was cloned at the multiple cloning site between the *Xho*I and *Bam*HI restriction sites.

Following linearization by *Eco*RI digestion, the pRS306[*PMS1*Δ] vector was transformed into yeast strain BGY8 (Section 3.3.13e). After 3 days incubation six Ura^+ colonies were observed. These colonies were screened for the presence of the integrating vector at the *PMS1* locus (Section 3.3.13f). Only one colony indicated the presence of the vector, producing a band ~5 k.b.p., the remaining colonies all maintained the wild-type *PMS1* locus and were presumably Ura^+ due to misintegration of the vector at another locus, or possibly gene conversion of the *ura3-52* locus. A few (~5) Ura^+ colonies were also obtained on the negative control

plate (no DNA transformed), a third possibility, therefore, is that the Ura^+ colonies represent mutant colonies that have spontaneously reverted to a Ura^+ phenotype.

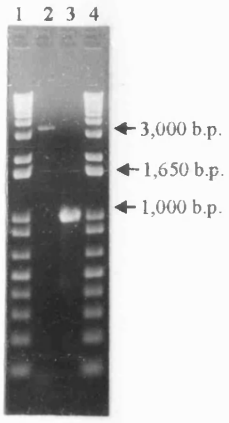


Figure 3.18 1% agarose gel showing the wild-type *PMS1* locus (Lane 2; 3382 b.p.), and the *PMS1* locus as amplified from strain BGY10, in which the *PMS1* ORF has been deleted (Lane 3; 982 b.p.). Lanes 1 and 4, the 1 k.b.p. DNA marker (Figure 2.1).

For the colony showing the presence of the vector, correct integration was further verified in the usual manner, by screening the novel junctions formed (Figure 3.14). Both the LHS and RHS screens produced the expected sized bands, 970 b.p. and 3342 b.p. respectively. The new strain containing the correctly integrated vector was designated BGY9 (Section 3.2.1).

Strain BGY9 was grown on 5-FOA medium in order to counter-select for the presence of the integrated vector (Section 3.3.13g). After 5 days incubation several colonies were observed growing, 10 of which were PCR screened. Recombination in 9 of the colonies had resulted in reversion to the wild-type locus, producing a 3382 b.p. band in the PCR screen (Lane 2; Figure 3.18). The remaining colony contained the desired chromosomal deletion, producing a band of 982 b.p., corresponding to loss of the *PMS1* ORF in this colony (Lane 3; Figure 3.18). The new *pms1*Δ colony was designated BGY10.

3.4.13 Construction of an I-*Sce*I expression vector

The expected sized fragment (720 b.p.) corresponding to the I-*Sce*I ORF was amplified from vector pPEX7 (Section 3.3.14b). By means of the *Eco*RI and *Xho*I restriction sites introduced at either end of the amplified fragment, the fragment was ligated into the yeast expression vector pYES2, and transformed by electroporation into *E. coli* (Section 3.3.14c). Ten ampicillin resistant colonies were PCR screened, 9 of the colonies contained vector with insert, each producing the desired 928 b.p. band. Plasmid DNA was prepared from one such colony and the sub-cloned fragment checked by DNA sequencing (Section 3.3.14d). Comparison of the sequence obtained form the sub-cloned DNA fragment with the universal code equivalent of the I-*Sce*I gene (Section 3.3.14a) produced a surprising result. It appeared at first that the cloned fragment was awash with mutations, containing 161 differences to the expected sequence. However, analysis of the amino acid sequence from both sequences revealed them to be translationally identical (except that two threonines at positions 123 and 156, changed to leucines in the original universal code equivalent [Colleaux *et al.*, 1986], were now once again threonines). The comparative analysis of the two DNA sequences (and their translation products) is shown in Figure 3.19.

The DNA sequence obtained is an alternative universal code equivalent of the I-*Sce*I ORF to the one expected [it had been presumed that the synthetic sequenced (unpublished) used by A. Thiery and B. Dujon in the construction of the vector pPEX7 (Section 3.2.2f) would represent the same sequence as constructed by Colleaux *et al.* (1986)]. Apart from the difference noted, the fragment cloned into vector pYES2 contained no other mutations, and was therefore deemed suitable for transformation into the experimental strains (see below). The new vector, pYES2[I-*Sce*I], was as for plasmid pYES2 (Section 3.2.2b) except that it contained the I-*Sce*I ORF cloned between the *Eco*RI and *Xho*I sites in the multiple cloning region.

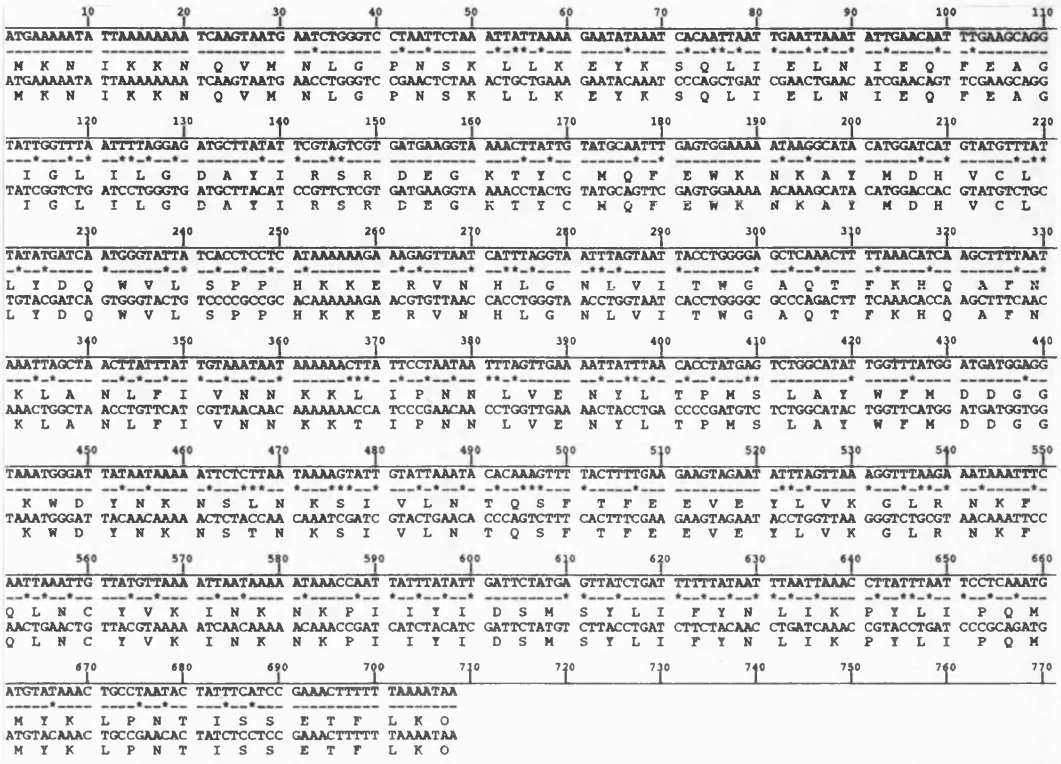


Figure 3.19 Comparative analysis of the universal code equivalent of the I-*Sce*I ORF as constructed by Colleaux *et al.* (1986) [upper sequence], with the DNA sequence obtained from the fragment sub-cloned from vector pPEX7 into pYES2 [lower sequence]. Base differences between the two sequences are indicated by (*). The translation product from both sequences are identical except for two leucine to threonine substitutions (see main text) at positions 123 and 156.

3.4.14 Transformation of the I-*Sce*I expression vector pYES2[I-*Sce*I] into the experimental strains BGY8 and BGY10

The experimental strains BGY8 and BGY10 were transformed by electroporation with the I-*Sce*I expression vector pYES2[I-*Sce*I], also with pYES2 only (Section 3.3.15). Following incubation several hundred Ura$^+$ colonies were observed growing for each transformation. One colony from each strain was PCR screened for the presence of the desired vector. In each case the expected band was produced; a band of 928 b.p. confirming the presence of the vector pYES2[I-*Sce*I], and a band of 245 b.p. confirming the presence of vector pYES2 only.

Strains BGY8 and BGY10, transformed with vector pYES2[I-*Sce*I] were designated as strains BGY11 and BGY12, respectively (Section 3.2.1). Strain BGY11 contains the completed experimental system, as described in Section 1.3. Strain BGY12 is the *PMS1*Δ derivative of BGY11.

Strains BGY8 and BGY10, transformed with vector pYES2 only, were designated as strains BGY13 and BGY14, respectively (Section 3.2.1). These strains are the negative control strains for BGY11 and BGY12, where no I-*Sce*I endonuclease expression can occur.

3.4.15 Construction of control strains

A fragment (1022 b.p.) containing the I-*Sce*I restriction site was amplified from a previously made DNA fragment (the *HML* gene disruption cassette) and ligated into the vector pRS313, and transformed into *E. coli* (Section 3.3.16a). Ten ampicillin resistant colonies were PCR screened, 5 of which contained vector with insert, producing the expected screening band of 1242 b.p. Plasmid DNA was prepared from one such colony and the I-*Sce*I restriction site verified by DNA sequencing (the concomitantly introduced *OR17-228* ORF was shown to contain the same mutations as noted in Section 3.4.8).

The new vector containing the I-*Sce*I restriction site was designated pRS313[I-*Sce*I]; the vector is as for vector pRS313 (Section 3.2.2h) except a fragment containing the I-*Sce*I restriction site has been cloned between the *Xba*I and *Xho*I sites in the multiple cloning region.

The first control strain - to show I-*Sce*I induced DNA cleavage - was constructed by transforming yeast strain FY833 with the vector pRS313[I-*Sce*I] and then with the I-*Sce*I expression vector pYES2[I-*Sce*I] (Section 3.3.16b). A second strain was transformed with pRS313[I-*Sce*I] and pYES2 only. The presence of the vectors was confirmed by PCR screening. The expected bands of 1242 b.p., 928 b.p. and 245 b.p. were obtained for pRS313[I-*Sce*I], pYES2[I-*Sce*I] and pYES2 only, respectively. The new strain containing vectors pRS313[I-*Sce*I] and pYES2[I-*Sce*I] was designated BGY17 (Section 3.2.1), the second strain containing vectors pRS313[I-*Sce*I] and pYES2 only was designated BGY18 (Section 3.2.1). In a third transformation the vector pPEX7 was transformed into FY833::pRS313[I-*Sce*I]. This

transformation was repeated twice, on both occasions only slow growing petite colonies were observed on the SC-His-Leu selective medium (see Discussion).

The second - gene conversion - control strain was constructed by transforming yeast strain BDY5 with the vector pYES2[I-*Sce*I] or pYES2 only (Section 3.3.15c). Once again, the presence of both vectors was confirmed by PCR screening, each vector producing the expected sized band, as given above. Strain BDY5 containing pYES2[I-*Sce*I] was designated BGY15 (Section 3.2.1). Strain BDY5 containing pYES2 only was designated BGY16 (Section 3.2.1). The nature of these 'gene conversion' control strains is discussed further in Section 3.5.1.

3.5 DISCUSSION

3.5.1 Generation of experimental strains for the analysis of molecular evolution by gene conversion

A) *Construction of experimental strains*

The disruption cassette for the incorporation of the *OR17-40* ORF at the *LEU2* locus, and an integrating vector for the incorporation of the *OR17-228* ORF at the *HML* locus, were both successfully made using routine recombinant methods. Similar strategies, involving counter-selection of a *URA3* marker, were successfully employed to leave both odorant receptor sequences integrated at their respective loci. Correct integration was confirmed by PCR screening of the novel joints created at either locus, and by DNA sequencing. Four mutations were noted in the *OR17-40* sequence (Table 3.13) and five in the *OR17-228* sequence (Table 3.14), resulting in a slight increase (0.4%) in overall homology between the two sequences. The viability of the I-*Sce*I sites at the new acceptor and donor loci was demonstrated, in each case, by successful digestions with I-*Sce*I enzyme.

The first method attempted to integrate the *OR17-228* ORF at the *HML* locus, direct gene replacement (Section 3.4.6), was unsuccessful. In this strategy a previously integrated *URA3* gene at the *HML* locus (in strain BGY3) was to be removed by a second transformation experiment, followed by selection on 5-FOA medium. It is unclear why this approach did not work. Possibilities include low transformation efficiency (or gene targeting efficiency) or insufficient pre-incubation of cells (to allow expression of a Ura^- phenotype) prior to plating on 5-FOA medium. It was expected that strain BGY3, when grown on medium lacking uracil, should exhibit temperature dependent growth (due to the *sir4-9* mutation; Section 3.3.3). Evidence of such temperature dependent growth was investigated using the closely related strain BGY4, but was not observed (Section 3.4.4). An increase in the distance between the *E* and *I* elements at *HML* can result in a lessening of *SIR* regulation (Schnell and Rine, 1986), however, in this instance, the introduced *URA3* gene resulted in the distance between the two silencing elements being reduced (by approximately 1.5 k.b.p.), thus ruling out this effect. A likely explanation for the absence of a temperature dependant growth effect, is that the onset of repression at *HML* (at the permissive temperature; 24°C), was prevented by the locus being

maintained in a transcriptionally active (unrepressed) state, due to strong induction (in the absence of uracil) of the *URA3* gene at the locus.

Although the effect of the *sir4-9* mutation was not apparent for *URA3* at *HML*, during the construction of the Gal^+ experimental strain (discussed below) its effect was observed; where certain colonies showing a haploid phenotype at 24°C (producing a 'halo' on the mating-type tester plate), appeared as diploids at 34°C (no 'halo'). The more pronounced effect may reflect better regulation (silenced/unsilenced state) of the native *HML* and *HMR* loci, or the nature of the effect being observed, i.e. whereas the *URA3* gene is normally regulated as part of a metabolic pathway (and was maintained in an induced state, as described above) the mating-type genes were responding to changing levels of their natural regulator: **a**1/α2 protein formation (Section 1.1.1).

In order to convert the experimental strain to a His^+ phenotype, strain BGY6 was transformed with a linear fragment containing the *HIS3* gene. This strategy was successful, resulting in numerous His^+ colonies, one of which was selected for the next step in the construction of the experimental strain. However, whilst being relatively simple to implement, this strategy is untidy in its outcome, in that the nature of the integration event cannot be easily determined. But it seems highly probable that the fragment would have been targeted to the intended *his3-532* locus of strain BGY6.

A Gal^+ experimental strain (BGY8) was successfully created by mating strain BGY7 with strain INVSC2, followed by sporulation and analysis (Section 3.4.10). Although the effect of the *sir4-9* allele had been revealed during the mating-type analysis (above), for the newly created experimental strain it was not possible to determine if the allele was present. Because of this uncertainty it was necessary to induce the experimental strains at the lower temperature of 24°C, to ensure silencing at the donor locus.

The galactose inducible nature of strain BGY8 was investigated by means of a β-galactosidase assay (Section 3.4.11). The expected range of values for the β-galactosidase activity is unknown. However, the assay was performed with the aim of comparing the newly created Gal^+ strain with the known galactose inducible strain FY10. The Gal^- parent strain (BGY7) was also compared. The assay itself was found to be more qualitative than quantitative in nature, and originally quite subjective in the

determination of the 'end-point' (defined as the time when the assay turns bright yellow). To remove this uncertainty each assay was stopped after a pre-determined amount of time (1 min 30 s). The highest activity, 0.25 - 0.27 $ml^{-1}min^{-1}$ was seen for the strain FY10Z (Table 3.16). β-galactosidase activity, although lower, 0.16 - 0.20 $ml^{-1}min^{-1}$, was also demonstrated for the new Gal^+ strain, BGY8Z. Seemingly, activity in the same range was also shown for BGY7Z, 0.19 - 0.20 $ml^{-1}min^{-1}$, however, this result may not be an accurate indicator, as the cells for this assay were recovered in a different fashion to the two other strains; thus for FY10Z and BGY8Z approximately 250 µl from the 10 ml induction culture was, in each case, sufficient to provide the required number of cells for the assay, whereas for BGY7Z it was necessary to centrifuge the entire induction culture (10 ml) to collect the required number of cells, i.e. BGY7Z exhibited no growth in galactose medium. Because of this clear difference in growth between BGY7Z and BGY8Z, and the demonstrated β-galactosidase activity shown by the BGY8Z cells, it was deemed satisfactory to proceed using strain BGY8 as the galactose inducible experimental strain.

A *PMS1*Δ derivative of the experimental strain was successfully created using the pop-in/pop-out gene replacement strategy (Section 3.3.3). The strain thus constructed, BGY10, was found to grow approximately 3 times slower than its parental strain, BGY8. When grown on solid medium a large variety of colony sizes was observed. The aberrant growth of BGY10 is most likely due to an accumulation of harmful mutations, as a result of the *PMS1* gene knock-out, which has been shown to have a mutator phenotype (Kramer *et al.*, 1989).

Strains BGY8 and BGY10 were transformed with the I-*Sce*I expression vector pYES2[I-*Sce*I], creating the completed experimental strains BGY11 and BGY12, respectively.

B) *Construction of control strain BGY15*

Strain BGY15, the control strain for the gene conversion process (no donor locus present), was created by transforming strain BDY5 (an FY-based strain) with the vector pYES2[I-*Sce*I]. Although BGY15 contains the desired attributes (Section 3.3.11), it would have been preferable to have constructed the strain using a JRY50 based derivative (as for the experimental strains) to allow a more valid comparison to be made [the order in which the experimental strains were constructed

was prohibitive to this task]. However, strain BGY15 is a satisfactory control strain for the preliminary experiment reported.

3.5.2 Conclusions

The strain required for the gene conversion experiments, containing the model system shown in Figure 1.19, was successfully made (strain BGY11). A mismatch repair defective derivative of this strain, containing a deletion of the *PMS1* open reading frame, was also successfully made (strain BGY12). Additionally, two 'vector-only' control strains for BGY11 and BGY12, namely BGY13 and BGY14, respectively, were also successfully made. Also described in this chapter is the construction of two types of control strains, for looking at various aspects of the experimental reaction: the 'vector control' strain BGY17 (and its 'vector-only' partner, BGY18); and the 'gene conversion' control strain BGY15 (and its 'vector-only' partner, BGY16).

The experiments performed with each of the strains constructed are described in the next chapter.

CHAPTER 4

DIRECTED MOLECULAR EVOLUTION BY GENE CONVERSION: INDUCTION OF EXPERIMENTAL AND CONTROL STRAINS

4.1 INTRODUCTION

This chapter describes the experiments performed with each of the strains - experimental and control - constructed in the previous chapter. A brief description of the experiments to be carried out is given below. The experimental protocol, for each experiment, is described in detail in the Methods section.

Experimental strains

Naturally, the aim of the experiment using the experimental strains is to demonstrate gene conversion at the acceptor locus, as described in the experimental strategy (Section 1.3). In this first experiment, gene conversion events are monitored by loss of a restriction site, *Hin*dIII, that is unique to the acceptor locus (the arrangement of the donor and acceptor loci is shown in Figure 1.19). The *Hin*dIII site is convenient as it lies towards the 3'-end of the *OR17-40* ORF (146 b.p. upstream from the termination codon) and is thus likely to be lost early on in the 'shuffling' process. The presence or absence of the *Hin*dIII restriction site following induction will be monitored by amplifying the acceptor locus by PCR, followed by restriction digest and analysis of the fragments produced by agarose gel electrophoresis.

Two principal experimental strains were constructed, BGY11 and BGY12, also constructed were 'vector-only' controls for both of these strains, BGY13 and BGY14. In the experiment described, only the principal experimental strain, BGY11 and its mismatch repair deficient derivative, BGY12 are used. For both strains

induction of shuffling is by 24 h growth in galactose medium (for both strains a control experiment is performed where induction is performed in glucose medium instead of galactose).

Control strains

'Vector control': The experiment using the 'vector control' strain BGY17 (and its 'vector-only' counterpart, BGY18) is designed to show I-*Sce*I activity by means of a vector-loss assay: induction of the endonuclease, by 24 h growth in galactose containing medium, should result in cleavage of the vector - pRS313[I-*Sce*I] - containing the I-*Sce*I recognition site. Subsequent loss of the vector (containing the *HIS3* auxotrophic marker) will be followed by the resulting growth deficiency on selective growth medium (for both strains a control experiment is performed were induction is carried out in glucose containing medium instead of galactose). The experiment is described following induction in liquid medium and on solid medium. Possible repair events following induction, e.g. non-homologous end-joining, are monitored by PCR screening (selected colonies).

'Gene conversion' control: The second control experiment, using the 'gene conversion' control strain BGY15 (and its 'vector-only' counterpart, BGY16) is designed to look at the kinetics of I-*Sce*I induced chromosomal double-strand cleavage. It is expected that a double-strand breakage, where no chromosomal homologue exists for repair, should result in a reduction in cell viability with time (Fairhead and Dujon, 1993). Induction is performed in galactose medium, and in glucose medium as a control. At various time points following induction (see Methods), cells are plated onto complete medium (glucose) in order to determine plating efficiency. In a second experiment, cells are plated onto complete medium containing galactose, again in order to look at plating efficiency. The cells for the second experiment are taken (at two time points) from the glucose induced control cultures for the first experiment. As for the 'vector control' experiment (above), possible repair events in the induced cells are monitored by PCR screening (selected colonies).

The experiments described above, for both the experimental and control strains, represent the initial experiments performed with each of the strains constructed (see Discussion).

4.2 METHODS: INDUCTION OF EXPERIMENTAL AND CONTROL STRAINS

4.2.1 Induction of experimental strains

A single colony of the experimental strain (experiment was performed with strains BGY11 and BGY12; Section 3.2.1) was picked from an SC-Ura plate (2% glucose) into 1 ml of SC-Ura (2% raffinose) and grown for 24 hours at 24°C [this induction temperature was a result of the unknown nature of the *sir4-9* allele; see Discussion], with shaking at 250 r.p.m. The cells were washed twice with 1 ml of water and resuspended in 1 ml of water. I-*Sce*I endonuclease production was then induced by inoculating 10 ml of fresh liquid SC-Ura (2% galactose) medium with 5×10^7 resuspended cells and incubating for 24 hours at 24°C, with shaking at 250 r.p.m. After this induction period, approximately 300 cells were spread on a YPD plate which was then incubated for 3 days at 24°C. A control experiment in which 5×10^7 resuspended cells were transferred to fresh liquid SC-Ura (2% glucose), medium instead of SC-Ura (2% galactose), with subsequent growth of cells on a YPD plate, was also performed.

Selected colonies (see Results) were chosen for analysis. From those colonies the acceptor loci were amplified via Expand™ PCRs (Section 2.2.12b) using the primers Leu2PF3/Leu2TB3 (Table 5.4). The cycling parameters for the reaction was: [94°C, 5 min; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min + 5 s/cycle) × 22; 72°C, 5 min]. PCR products were analysed by agarose gel electrophoresis (Section 2.2.4) then purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c). Cleaned PCR fragments (~0.1 µg) were then digested (Section 2.2.14) with *Hin*dIII (10 units), at 37°C for 4 hours. Digested fragments were analysed by agarose gel electrophoresis (Section 2.2.4).

4.2.2 Induction of control strains

A) *Vector loss experiment*

A single colony of BGY17 (Section 3.2.1; also performed with BGY18) was picked from an SC-Ura-His plate (2% glucose) into 1 ml of SC-Ura-His (2% raffinose) and grown for 24 hours at 30°C, with shaking at 800 r.p.m. The cells were washed twice

with 1 ml of water and resuspended in 1 ml of water. I-*Sce*I endonuclease production was then induced by inoculating 10 ml of fresh liquid SC-Ura (2% galactose) medium with 5×10^7 resuspended cells and incubating for 24 hours at 30°C, with shaking at 250 r.p.m. After this induction period, approximately 300 cells were spread on a SC-Ura (2% glucose) plate which was then incubated for 3 days at 30°C. Vector loss was determined by picking the resultant colonies (100) onto SC-Ura-His (2% glucose) medium. The transferred colonies were incubated at 30°C for 2 days and inspected for growth. A control experiment in which 5×10^7 resuspended cells were transferred to fresh liquid SC-Ura-His (2% glucose) medium instead of SC-Ura-His (2% galactose), with subsequent growth of cells on a SC-Ura (2% glucose) plate and transfer of a 100 colonies to an SC-Ura-His (2% glucose) plate, was also performed.

The above experiment was also performed where the liquid induction of the cells was substituted for induction on solid medium. Hence, ~300 resuspended cells (as above) were inoculated onto a SC-Ura (2% galactose) plate and incubated at 30°C for 3 days. Resultant colonies were then picked onto SC-Ura-His (2% glucose) medium, as above. As for the liquid induction, a control induction was performed where the resuspended cells were inoculated onto a SC-Ura (2% glucose) plate, instead of a SC-Ura (2% galactose) plate, with subsequent transfer of cells onto SC-Ura-His (2% glucose) medium, as above.

Selected colonies (see Results) were chosen for analysis. From those colonies vector sequences (containing the I-*Sce*I recognition site) were amplified via *Taq* PCRs (Section 2.2.12b) using the primers T7/M13 (Table 5.2). The cycling parameters for the reaction was: [94°C, 5 min; (94°C, 30 s; 48°C, 30 s; 72°C, 1 min 30 s) × 30]. PCR products were analysed by agarose gel electrophoresis (Section 2.2.4) then purified using the Wizard™ PCR Preps DNA Purification Kit (Section 2.2.4c). Cleaned PCR fragments (~0.1 μg) were then digested (Section 2.2.14) with I-*Sce*I (5 units), at 37°C for 2 hours. Digested fragments were analysed by agarose gel electrophoresis (Section 2.2.4).

B) *Reduction in viability following a chromosomal DSB*

A single colony of BGY15 (Section 3.2.1; also performed with BGY16) was picked from an SC-Ura plate (2% glucose) into 1 ml of SC-Ura (2% raffinose) and grown for 24 hours at 30°C, with shaking at 800 r.p.m. The cells were washed, and I-*Sce*I

expression induced as described above (Section 4.2.1). At time intervals of 2, 3, 4, 6, 8, 10 and 24 hours, approximately 300 cells (Section 2.2.11) from the induction culture were inoculated onto two YPD plates (Section 2.2.7), and grown for 3 days at 30°C. A control experiment in which the resuspended cells were transferred to fresh liquid SC-Ura (2% glucose), medium instead of SC-Ura (2% galactose), with subsequent plating onto YPD medium (at the above time intervals) was also performed. For the control assay, at the 10 and 24 hour time points, approximately 300 cell were also plated onto YPG medium (Section 3.3.11e).

Selected colonies (see Results) were chosen for PCR analysis. From those colonies the acceptor locus (although there is no donor locus in this strain) was amplified and purified as described above (Section 4.2.1). The cleaned PCR products were then digested with I-*Sce*I, as described above (Section 4.2.2a).

4.3 RESULTS: INDUCTION OF EXPERIMENTAL AND CONTROL STRAINS

4.3.1 Induction of experimental strains

The experimental strains BGY11 and BGY12 were induced as described in Section 4.2.1. The acceptor locus from 5 galactose and 5 glucose induced colonies (for each strain) growing on YPD medium were amplified and analysed by agarose gel electrophoresis. Each colony produced a band of the expected size: 1847 b.p. The PCR products were purified, then digested with the restriction enzyme *Hin*dIII to look for evidence of gene conversion (loss of the restriction site). Each sample, for both strains and both induction conditions (galactose/glucose), produced the digest fragments corresponding to the presence of the *Hin*dIII site, i.e. no evidence of gene conversion at the *Hin*dIII site was observed. The acceptor loci amplified from BGY11 colonies, and the fragments obtained following *Hin*dIII digestion of the amplified fragments, are shown in Figure 4.1. The same pattern was observed for BGY12.

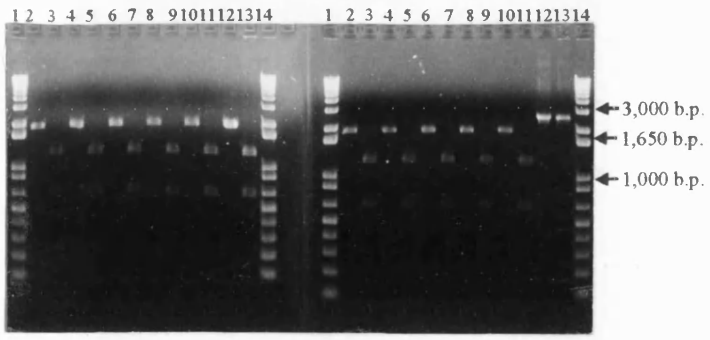


Figure 4.1 1% agarose gel showing screening bands for the induced experimental strain BGY11. The bands at 1847 b.p. in the LHS gel represent the acceptor loci amplified from 5 galactose induced BGY11 colonies. Next to each acceptor locus band are the expected bands (1207 b.p. and 640 b.p.) obtained following *Hin*dIII digestion of each fragment. The RHS gel shows the same PCR screen and subsequent *Hin*dIII digestion for 5 glucose induced BGY11 colonies. Lanes 1 and 14 (both gels), the 1 k.b.p. DNA marker (Figure 2.1). [Bands in lanes 12 and 13 (RHS gel) represent controls for the PCR amplification].

4.3.2 Induction of control strains

A) *Vector loss experiment*

The control strains BGY17 and BGY18 were induced as described in Section 4.2.2a. For each strain 100 galactose and 100 glucose induced colonies were picked onto SC-Ura-His selective medium and inspected for growth. Induction was performed on both liquid medium and solid medium. The percentage of colonies growing for each experiment is shown in Table 4.1; discussed in Section 4.4.1.

Strain	Galactose (liquid)	Glucose (liquid)	Galactose (solid)	Glucose(solid)
BGY17	50%	61%	76%	71%
BGY18	93%	85%	92%	93%

Table 4.1 The percentage of colonies growing on SC-Ura-His selective medium following induction in liquid and on solid medium. Each figure represents the combined results of two experiments.

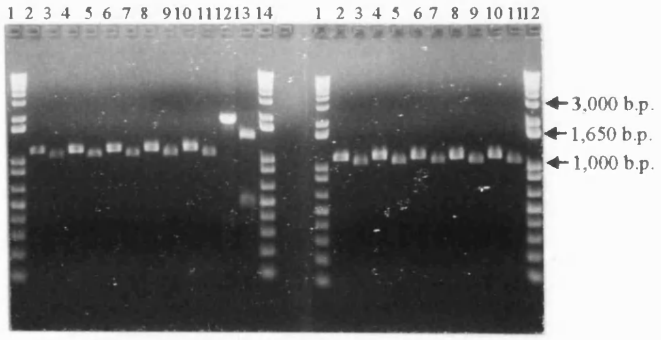


Figure 4.2 1% agarose gel showing screening bands for the induced control strain BGY17. The bands at 1242 b.p. in the LHS gel represent the vector sequence amplified from 5 solid (galactose) induced BGY17 colonies. Next to each band is the expected fragment (1107 b.p.) obtained following I-*Sce*I digestion (the smaller 135 b.p. digests fragment is not observed). [Bands in lanes 12 and 13 represent a control I-*Sce*I digestion of known DNA]. The RHS gel shows the same PCR screen and subsequent I-*Sce*I digestion for 5 solid (glucose) induced BGY17 colonies. Lanes 1 and 14 (LHS gel) and 1 and 12 (RHS gel), the 1 k.b.p. DNA marker (Figure 2.1).

Five BGY17 colonies growing on the SC-Ura-His selective medium, from each of the induction procedures (liquid galactose/glucose, solid galactose/glucose), were analysed for evidence of vector repair (change in expected PCR product/loss of I-*Sce*I recognition site). Vector sequences were amplified by PCR and analysed by agarose gel electrophoresis. Each colony produced a band of 1242 b.p., corresponding to the original, unmodified, vector sequence. The PCR products were purified and the I-*Sce*I recognition site checked by digestion. The expected digestion bands (following cleavage at an intact I-*Sce*I site) were observed for each sample. Figure 4.2 shows the amplified vector sequences from the five solid (galactose) and five solid (glucose) induced BGY17 colonies, and the fragments obtained following I-*Sce*I digestion. The same pattern was observed for the liquid induced colonies.

B) *Reduction in viability following a chromosomal DSB*

The control strains BGY15 and BGY16 were induced as described in Section 4.2.2b. At each time point approximately 300 cells were transferred from the induction medium (glucose and galactose induction done for both strains) and inoculated onto two YPD plates, where they were incubated at 30°C for 3 days. The plates were then inspected for colony formation. At every time point, for both strains, and both induction conditions (galactose/glucose), the same number of colonies were observed growing on each YPD plate; no reduction in cell viability was observed. Five galactose and 5 glucose, 24 hour induced BGY15 colonies were screened for evidence of repair at the acceptor locus (change in expected PCR product/loss of I-*Sce*I recognition site). The acceptor loci were amplified by PCR and analysed by agarose gel electrophoresis. Each colony produced a band of 2036 b.p., corresponding to an unmodified acceptor locus. The PCR products were purified and the I-*Sce*I recognition site checked by digestion. The expected digestion bands (following cleavage at an intact I-*Sce*I site) were observed for each sample. Figure 4.3 shows the acceptor loci fragments amplified from 24 hour galactose induced BGY15 colonies, and the fragments obtained following I-*Sce*I digestion. The same pattern was observed for the 24 hours glucose induced colonies.

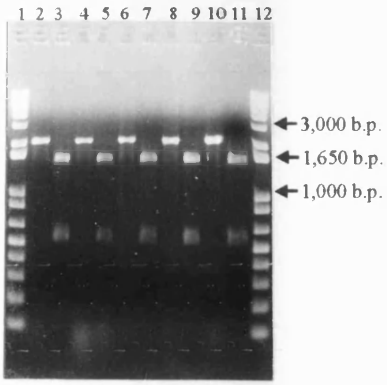


Figure 4.3 1% agarose gel showing the screening bands obtained from the induced control strain BGY15. Bands at 2036 b.p. represent the acceptor loci amplified from 24 hour galactose induced BGY15 colonies. Next to each band are the expected fragments (540 b.p. and 1496 b.p.) obtained following I-*Sce*I digestion. Lanes 1 and 12, 1 k.b.p. DNA marker (Figure 2.1).

Cells grown in glucose induction medium (as well as being inoculated onto YPD plates, above) were also inoculated onto galactose containing YPG plates (300 cell/plate), in order to look for a reduction in viability. Plating onto YPG medium was performed at two (convenient) time points, 10 and 24 hours. Plated cells were incubated at 30°C for 5 days, and the number of colonies determined. The results are presented in Table 4.2; discussed in Section 4.4.1.

Strain		Galactose (solid)	Glucose (solid)
BGY15	10 hours	160	316
	24 hours	140	269
BGY16	10 hours	221	283
	24 hours	252	287

Table 4.2 The number of colonies of BGY15 and BGY16 growing on galactose and glucose containing solid medium. The expected number of colonies per plate is 300. Each number represents the average number of colonies from two plates.

The acceptor locus from five BGY15 colonies growing on a YPG plate and from five growing on a YPD plate were analysed by PCR, with subsequent digestion of the amplified fragments with I-*Sce*I. As for the 24 hour liquid induced BGY15

colonies (above), no differences were observed in the size of the fragments amplified from the colonies grown on galactose containing medium compared with those grown on glucose containing medium. Each colony producing a band of 2036 b.p. Digestion revealed the I-*Sce*I site to be intact in the amplified fragments. The same pattern of amplified fragments and digestion products as shown for 24 hour liquid induced BGY15 colonies (Figure 4.3), was again observed.

The results presented above for both the experimental and the control strains are discussed in the next Section.

4.4 DISCUSSION

4.4.1 Induction of experimental and control strains

The experimental strains BGY11 and BGY12 were induced for I-*Sce*I expression by growth in liquid galactose-containing medium for 24 hours. The induced cells were then grown on YPD medium and evidence of gene conversion events determined by PCR screening and restriction digestion (Section 4.3.1). Five galactose and five glucose induced colonies were screened for both strains. Disappointingly, no differences were observed between the cells grown in either medium; for both strains the *Hin*dIII restriction site remained intact, indicating that the intended gene conversion reaction had not taken place.

It seems likely, however, that failure of either of the experimental strains to perform the intended reaction is not due (in this instance) to a flaw in the experimental strategy, but rather due to failure of one of the components of the experimental system, namely I-*Sce*I expression, as demonstrated by both control experiments (below).

In the first control experiment, evidence of vector cleavage following I-*Sce*I expression was examined, with induction performed in liquid and on solid medium (Section 4.3.2a). In the second control experiment, cleavage of a chromosomal I-*Sce*I site was examined, again induction was performed on solid and in liquid medium (Section 4.3.2b). For both experiments, growth analysis (Table 4.1, 4.2 and Section 4.3.2b) revealed no (significant) difference between galactose and glucose induced cells (or between the second set of control strains carrying the pYES2 only vector). Further, PCR screening of selected colonies (both experiments) revealed the I-*Sce*I recognition site to be intact, and there to be no change in the expected PCR product, i.e. no evidence of repair (vector or chromosome). This strongly indicated that I-*Sce*I expression from the constructed expression vector, pYES2[I-*Sce*I], was not occurring as intended.

It would have been particularly useful to have been able to perform either of the control experiments using the original I-*Sce*I expression vector, pPEX7. Several attempts were made to construct a 'vector control' strain using pPEX7, where the vector was transformed into FY833::pRS313[I-*Sce*I] with subsequent selection on SC-His-Leu medium (Section 3.3.16b). On every occasion no viable colonies were produced (pPEX7 was also transformed into FY833 only, with selection on SC-Leu

medium, with the same result). It is unclear why these transformations failed. Possibly pPEX7 (carrying the *LEU2d* marker) failed to complement the *leu2Δ1* allele found in FY833 [although this explanation is inconsistent with the observations of Fairhead and Dujon (1993), where pPEX7 was successfully used in another FY-strain, FY23, carrying the same *leu2Δ1* allele]. No attempt was made to transform pPEX7 into an alternative strain (discussed in Chapter 6).

Future work

Naturally, the immediate aim of any future work must be to establish whether poor endonuclease expression from vector pYES2[I-*Sce*I] (upon galactose induction) is the cause of the unsatisfactory results reported. For the experimental strategy to succeed it is imperative that I-*Sce*I be expressed correctly and in a tightly controlled manner. A potential first step in this investigation would be to look at the expression profile by means of a Northern blot.

4.4.2 Conclusions

The results presented in Section 4.3 represent the initial experiments performed with each of the strains constructed (experimental and control). A great deal more work is needed, firstly to rectify the immediate problem of poor I-*Sce*I expression, then later to investigate the ability of the experimental systems to perform the intended gene shuffling reaction. The limited nature of the results to date, and a general perspective on the construction of the experimental strains, is discussed in Chapter 6.

CHAPTER 5

TOWARDS A CRE-BASED RECOMBINATION SYSTEM FOR GENERATING INTEGRATED DNA REPERTOIRES SITE-SPECIFICALLY IN YEAST

5.1 INTRODUCTION

The Cre-*lox* recombination system is a recent addition to the ever expanding tool-box of the molecular biologist, allowing site-directed recombination to be induced *in vitro*, or *in vivo*, in a variety of organisms.

The system has been used in yeast to delete genes and to epitope tag them. It has also been used to integrate - via a single cross-over - exogenous DNA elements into the yeast genome. As will be seen, a drawback of the last application is that plasmid DNA is cointegrated with the exogenous element. The aim of the work presented in this chapter is to investigate the efficiency of a more precise Cre-mediated integration reaction, one based on a double recombination event (described in detail later). It is envisaged that the double cross-over reaction will allow exogenous DNA elements to be integrated site-specifically into the yeast genome with greater efficiency than is possible with current gene replacement methods (discussed later).

A brief review of the Cre-*lox* recombination system is given below. This is followed by details of the overall experimental strategy.

5.1.1 The Cre-*lox* site-specific recombination system

The site specific recombination system of bacteriophage P1 was first characterised in 1981 by N. Sternberg and D. Hamilton (Sternberg and Hamilton, 1981). A 6.5 k.b.p. *Eco*RI fragment (designated as fragment 7) of P1 DNA was cloned into a λ vector. This fragment was shown to efficiently reassort λ markers flanking the fragment in

bacteria lacking *rec*A and *rec*BC functions. The study identified two components necessary for this recombination activity, a site called *loxP* (for *lo*cus of crossing-over (*x*), *P*1), and a *trans*-acting function called *cre* (for *c*auses *re*combination). The *loxP* site and the *cre* gene were found to be confined to separate, but adjacent regions within a 3 k.b.p. segments of the *Eco*RI-7 DNA fragment. In a second paper (Sternberg *et al.*, 1981), the efficiency of recombination between a λ-P1 hybrid phage (λP1:7) and the bacterial chromosome was investigated. It was found that integrative recombination between λ-P1:7(*loxP*) DNA and the bacterial chromosome (at a site called *loxB*) resulted in an integrated λ prophage, flanked by two hybrid *lox* sites: *loxL* and *loxR* (Figure 5.1). The recombination potential of all four sites (*P*,*L*,*R* and *B*) was determined. The crosses revealed *P*×*B*, *L*×*R* and *P*×*R* to be inefficient in recombination, whilst *L*×*L* and *P*×*L* were as efficient as *P*×*P* recombination. The recombination frequency between the *loxP* and *loxB* sites was found to be very low; this suggested that prophage formation did not play a significant role in the life cycle of P1.

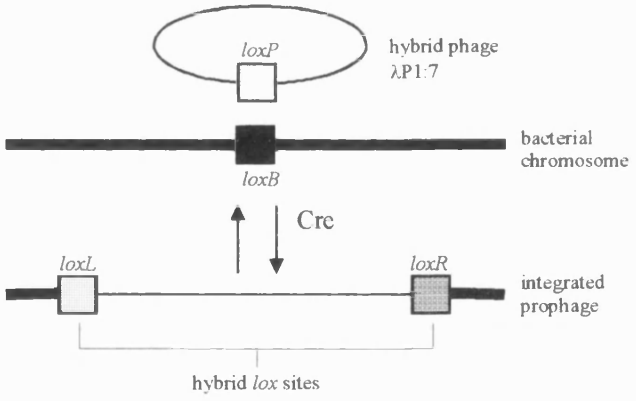


Figure 5.1 Integrative recombination between the hybrid phage λP1:7 and the bacterial chromosome.

Thus, unlike other phage (such as λ) the recombination machinery of P1 is not required for integration into the genome of its host (*E. coli*). Instead, the virus exists as a circular DNA molecule maintained at one to two copies per cell (Ikeda and Tomizawa, 1968). Sternberg *et al.* (1981), showed that *loxP* recombination could promote (but was not required for) circularization of injected DNA. However, the primary role of the recombination system in P1 was demonstrated to be in the stable maintenance of the P1 plasmid: during replication, pairs of homologous DNA

molecules are created, between which recombination can occur by exchange of sister DNA strands. The consequences of such a reciprocal recombination event for a unit copy replicon, such as P1, is the formation of a replicon dimer. A single dimer in a cell about to divide cannot be partitioned before cell division, resulting in one daughter cell that lacks the plasmid. In *E. coli* cells that harbour P1, fewer than one in 10^4 cell divisions produces a cured cell. Austin *et al.* (1981), showed that this fidelity was due to the Cre-*loxP* recombination system. The system promotes rapid recombination between two *loxP* sites on dimer molecules, resolving them into monomeric substrates for proper partition. It was found that intramolecular recombination was more efficient than the intermolecular event, promoting dimer breakdown faster than dimer production. This difference in efficiency between the two recombination reactions has also been noted *in vitro* (Abremski and Hoess, 1984).

The sequence of the *loxP* site (also the *loxB*, *loxL* and *loxR* sites) was determined by Hoess *et al.* (1982). The 34 b.p. site is composed of two perfect 13 b.p. inverted repeats, separated by an asymmetric 8 b.p. spacer region (Figure 5.2). The *loxB* site was found to be an imperfect version of this site, having 65% homology to the *loxP* site within the inverted repeats, and 50% homology in the spacer region. The main difference was the addition of an extra residue in the spacer region, making it 9 b.p. long (Figure 5.2).

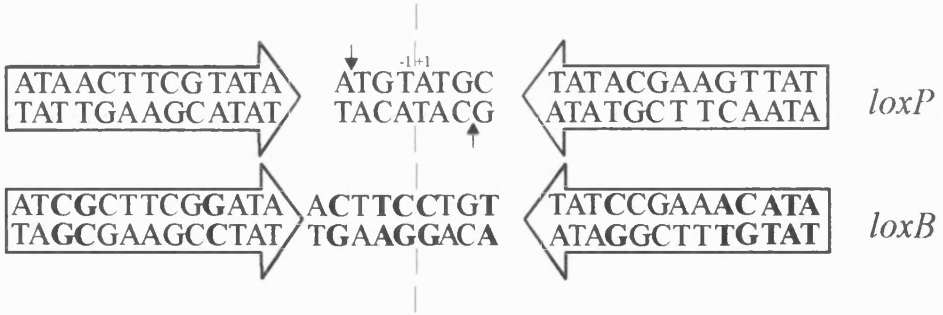


Figure 5.2 DNA sequences of the *loxP* and *loxB* sites. The inverted repeats are enclosed within the large arrows and separated by the spacer region. The points at which Cre cleaves the *loxP* site during recombination are denoted by the small vertical arrows. The positions of the bases with respect to the axis of dyad symmetry are numbered. Differences between the two sites are highlighted in bold.

Analysis of the *loxL* and *loxR* sites (formed by *loxP*×*loxB* recombination), revealed that the sites were, as expected, hybrids of *loxP* and *loxB*. An important

observation was that the cross-over point was within the spacer region (although undefined at this point). The differences noted in the recombination potential of the four sites (discussed above), was attributed to differences in the spacer region: *loxL*, having an 8 b.p. spacer like *loxP*, is proficient in recombination, whilst *loxB* and *loxR*, both containing a 9 b.p. spacer, are inefficient. By cloning the *loxP* site into a plasmid and constructing a series of deletion mutants, it was shown that all the information required for efficient Cre mediated recombination was located within a 63 b.p. sequence (+23 to -40; numbering as Figure 5.2) that flank the region of dyad symmetry. However, by means of a nuclease protection assay, it was later shown by Hoess and Abremski (1984) that only the 34 b.p. Cre recognition site was required. This study also revealed that the *lox* sites serve as a DNA binding domain for Cre proteins, each inverted repeat, and 4 b.p. of the contiguous adjoining spacer region, being bound (and protected from nuclease degradation) by Cre. The absolute stoichiometry was confirmed to be one molecule of Cre bound per inverted repeat in a later study by Mach *et al.* (1992).

Comparison of the DNA sequence of the *loxP* site (above) with the recognition site of other site-specific recombinases reveals an evolutionary relationship. The *FRT* site, recognised by the yeast 2-μm FLP recombinase, has an identical organisation to that of *loxP*, i.e. two 13 b.p. repeats (homologous to those of *loxP*), separated by an 8 b.p. spacer region (Senecof *et al.*, 1985). Homology is also seen (to a lesser extent) between the inverted repeat of *loxP* and the *att* site of the λ integrase.

The Cre recombinase gene was isolated by Sternberg *et al.* (1986). Using the *Eco*RI-7 DNA fragment from the original study (the fragment shown to reassort markers in a λ vector), a series of deletion mutants were created and screened for Cre activity. As a result, a 1.1 k.b.p. DNA sequence was localised which encoded a 343 amino acid (38 kDa) protein.

The requirements of the Cre-*lox* system have been shown to be very simple (Abremski *et al.*, 1983). Requiring only the Cre protein, Mg^{2+}, and no other cofactors, the system can successfully catalyse recombination *in vitro*. Furthermore, the conformation of the DNA substrate containing the *lox* site is not important; recombination occurs *in vitro* between *lox* sites on supercoiled, nicked circular or linear DNA molecules. In addition, the intramolecular *loxP*×*loxP* reaction occurs efficiently regardless of the orientation of the two sites relative to one another.

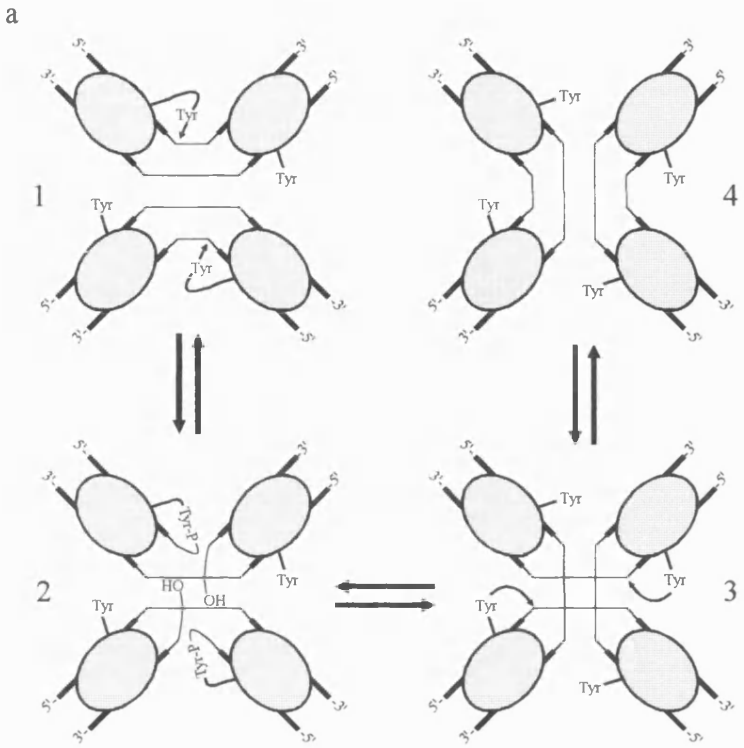


Figure 5.3a Mechanism of the Cre-*loxP* recombination reaction. Two Cre molecules bind to *loxP* sites which then associate to form a recombination synapse (**1**). Tyr-324 cleaves each substrate to form a covalent 3'-phosphotyrosine intermediate in an antiparallel arrangement (**2**). (Cleavage of the substrates could precede synapsis.) The 5'-hydroxyl groups released by cleavage of the phosphodiester linkages become nucleophiles in a strand transfer reaction, in which they attach the phosphotyrosines of the partner substrates, yielding a Holliday-junction intermediate (**3**). A second round of cleavage and strand exchange reactions gives recombinant products (**4**). The reaction intermediate shown in **2** represents the Cre-*loxA* synapse structure crystallised by Guo *et al.* (1997). Diagram reproduced from the latter study.

Sequence analysis of the *loxL* and *loxR* sites (generated upon integration of P1 into the *loxB* site of *E. coli*) had shown that the cross-over point during recombination was located somewhere within the 8 b.p. spacer region. The exact position of cleavage and strand exchange during recombination was determined by Hoess and Abremski (1985). By systematically replacing the phosphates in the DNA backbone of the spacer region with ^{32}P, and studying the products formed by recombination between a labelled and an unlabeled *loxP* site, the point of strand exchange during recombination was located (this was shown to coincide with the bond cleaved). It was found that Cre makes a 6 b.p. staggered cut within the *loxP* spacer during

recombination (Figure 5.2), resulting in a 5' protruding terminus. It was also shown that during strand exchange Cre becomes covalently attached to its DNA substrate via a phosphate, producing a free 5'OH. The nature of the covalent intermediate was revealed by Guo *et al.* (1997), who determined the three dimensional crystal structure (2.4 Å resolution) of the Cre protein bound to a *loxP* site. In this elegant study a 'suicide' *loxP* substrate (termed *loxA*) was used that halted the recombination reaction after strand cleavage, allowing an intermediate in the recombination reaction to be observed. The crystal structure of the trapped recombination complex revealed that four recombinases, and two *loxP* sites, form an antiparallel synapsed structure. Within this structure the DNA resembles a four-way Holliday-junction intermediate. The covalent linkage between the Cre protein and its DNA substrate is via a 3'-phosphotyrosine linkage, formed when a catalytic tyrosine residue (Tyr-324) cleaves its DNA substrate. The mechanism proposed by Guo *et al.* for strand exchange within the Cre recombination complex, and the *modus operandi* of the suicide *loxA* substrate are described in Figure 5.3a:b.

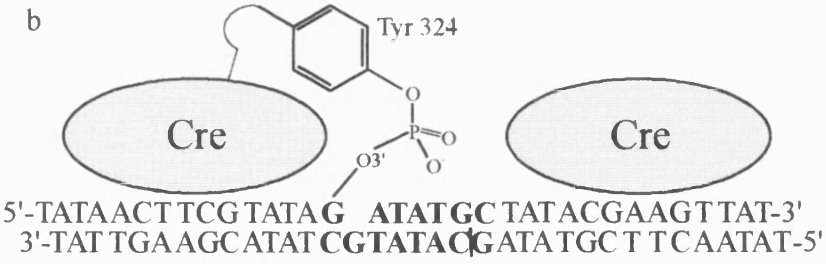


Figure 5.3b The trapped 3' phosphotyrosine intermediate formed when Cre cleaves the *loxA* substrate. (The *loxA* substrate was formed by annealing two modular *loxP* half sites through 4 b.p. 5' overhangs). The Cytidine3 residue on the 3' side of the cleaved phosphate is able to diffuse away, eliminating the 5'-hydroxyl group that would normally serve as a nucleophile in the subsequent strand exchange step of the reaction. Diagram reproduced from Guo *et al.* (1997).

The catalytic Tyr residue of the Cre protein, and three other residues required for catalysis (Arg-His-Arg), are strictly conserved amongst the >60 members of the integrase family of recombinases. The members of the family [which includes FLP (yeast) and Int (λ)] are characterised by their ability to carry out a complete site-specific recombinase reaction between two DNAs in the absence of any cofactors. This is achieved through the transient covalent recombinase-DNA linkage, which

conserves the high energy of the broken phosphodiester bond (Craig, 1988; Kwon *et al.*, 1997).

The recombination reaction catalysed by Cre exhibits directionality (Abremski *et al.*, 1983; Hoess and Abremski, 1984). When two *loxP* sites on the same DNA molecule are directly repeated, recombination results in the excision of the DNA between the sites. However, if the sites are in an inverted orientation with respect to each other, then upon recombination, the DNA between the sites undergoes inversion rather than excision. These events are shown diagrammatically in Figure 5.4.

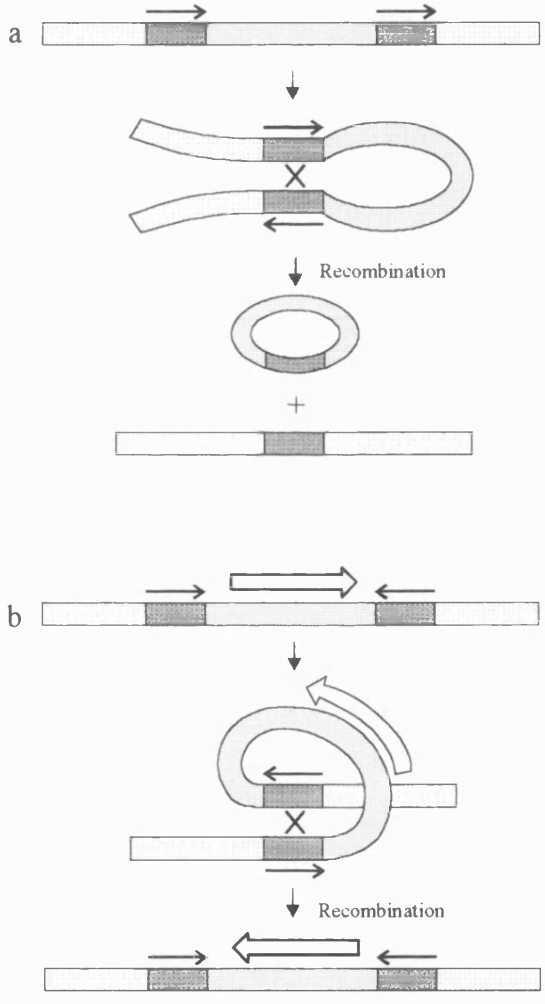


Figure 5.4 Excision and inversion promoted by Cre recombination. **a**. When two *lox* sites (dark rectangles) are directly repeated (arrows), recombination results in excision of the intervening DNA. **b**. When the *lox* sites are in opposite orientation, the intervening DNA element is inverted upon recombination (large arrow). Note that during synapsis (middle figures **a** and **b**) the *lox* sites have an antiparallel arrangement with respect to each other.

The *loxP* site is symmetrical except for the 8 b.p. spacer region. How does the asymmetry of the spacer region confer directionality on the site? Is the directionality achieved by DNA-DNA homology between the two recombining sites? These questions were addressed by Hoess *et al.* (1986) who constructed a series of mutant *loxP* sites containing single base pair alterations in the spacer region. Recombination analysis *in vivo* and *in vitro* revealed that homology was required between recombining *loxP* sites. Mutant sites (with changes at the +2 and +3 positions; numbering as Figure 5.2) were shown to recombine efficiently with themselves but not with a wild-type site. The study also revealed that the +1 position was critical for recombination. If the mutation was at this position, homology was found not to be sufficient, abolishing any recombination. The +1 position and its mirror image across the dyad axis of symmetry -1, form the dinucleotide TpA. A likely explanation, proposed for the role of the +1 site, is that the TpA sequence (which is predisposed to melting) is critical for unwinding of the *loxP* site, either during the pairing or the strand exchange step of recombination.

Cre recombinase in eukaryotes

Unlike the genome of prokaryotic cells, which is usually a single circular chromosome, the eukaryotic genome is typically larger, composed of multiple linear chromosomes, and is structurally more complex. A distinctive eukaryotic feature is that the DNA is organised into nucleosomes by intimate association with histones and other proteins to form chromatin. The ability of the Cre recombinase to act on a eukaryotic genome was first demonstrated by Sauer (1987). A selectable marker (*LEU2*) flanked by *loxP* sites was integrated into the yeast genome at two locations and in different orientations. Cre was induced from a vector under the control of the yeast *GAL1* promoter. The production of functional Cre was assayed by the appearance of leucine auxotrophs, which were shown to arise by excision of the *LEU2* marker. After 24 h induction, 98% of cells in the induced culture had a Leu$^-$ phenotype. In a subsequent study, Sauer showed that Cre was also functional (and non-toxic) in cells from a higher eukaryote, the mouse, effectively promoting recombination between *lox* sites placed on an introduced vector or between sites integrated at a chromosomal locus (Sauer and Henderson, 1988; Sauer and Henderson, 1989).

Because Cre normally acts in bacteria, it is unlikely to have any localisation signals to direct it to a eukaryotic nucleus. The functional radius of the nuclear pore has been measured to be 45 - 55 Å, sufficient to allow the passage of a spherical protein as large as 50 - 60 kDa (Barnes and Rine, 1985). Cre has a molecular weight of 38 kDa, and its ability to enter the nucleus may therefore be a consequence of its size (Sauer, 1987; Le *et al.*, 1999).

The ability of the Cre recombinase to function in eukaryotic organisms, combined with the overall simplicity of the system, marked the way for its use as a tool for genetic analysis and chromosome engineering.

DNA manipulation by site-specific recombination

Applications of the Cre-*lox* system have so far included: the creation of site-directed chromosome translocations in plants and mice (Qin *et al.*, 1994; van Deursen *et al*, 1995), the engineering of synthetic antibody libraries in phage (Waterhouse *et al.*, 1993; Tsurushita *et al.*, 1996), recombination activated gene expression (RAGE) in mice (Lasko *et al.*, 1992; Sauer, 1993), and the development of efficient multipurpose transposons, and disruption cassettes, for repeated gene disruption and epitope tagging in yeast (Hoekstra *et al.*, 1991; Ross-MacDonald *et al.*, 1997; Güldener *et al.*, 1996; Delneri *et al.*, 2000; Antoni and Gallwitz, 2000).

The examples presented above are based on the excision or rearrangement of DNA, caused by Cre induced recombination between chromosomally integrated *lox* sites. The strategy for generating antibody libraries in phage was slightly different, being an adaptation of the natural role of the Cre-*lox* system, whereby it is used to resolve a plasmid cointegrate, resulting in a more diverse library. However, of most interest to our own work is the use of the Cre-*lox* system as a means of introducing DNA constructs site-specifically into the genome.

Sauer (1996) showed that Cre could carry out such integrative recombination in yeast cells. A *leu2* auxotrophic strain was engineered to contain a single *loxP* site on chromosome VII, near *PDR1* (a permeability regulatory gene). It was shown that transformation with a *LEU2* targeting vector, bearing a single *loxP* site (with concomitant expression of Cre from a *GAL1-cre* expression vector), resulted in Leu^+ colonies via site-directed integration of the vector at the *PDR1* locus. Integration had occurred via a single intermolecular cross-over between the vector borne and chromosomal *lox* sites (Figure 5.5).

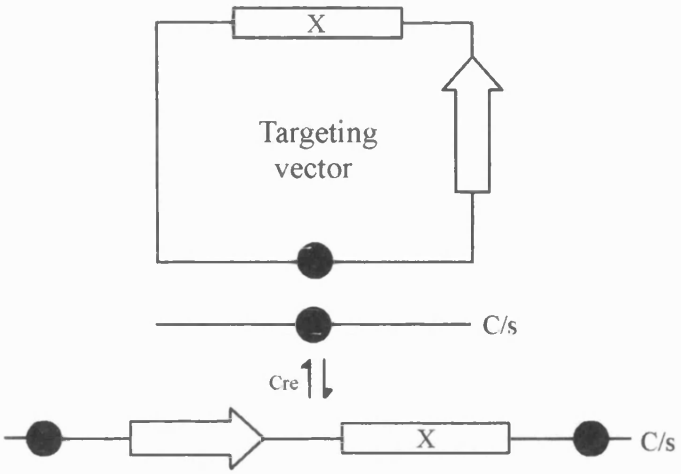


Figure 5.5 Cre-mediated site-specific integration of exogenous DNA into the yeast genome via a single cross-over event between identical *lox* sites (filled circle). X: DNA segment for insertion. C/s: chromosome. Arrow: nutritional marker gene, e.g. *LEU2*.

In the same study Sauer also investigated the ability of the Cre recombinase to target a cryptic *lox* site identified in the yeast genome. The *lox*-like site was located upstream of the *FAS1* gene (encoding the β subunit of fatty acid synthetase), giving the site its name, *loxFAS1*. The site was shown to have 9 b.p. of identity adjacent to the *loxP* core on one side, and a single mismatch in the 10 b.p. adjacent to the core on the other, with the core itself bearing almost no homology to the *loxP* core. The targeting vector was modified to contain the *loxFAS1* core (and two functional inverted repeats), producing a new site, *mock loxFAS1* (Figure 5.6). Sauer demonstrated that Cre targeting of the endogenous *FAS1* locus with the *mock lox* vector was only 2-fold less efficient than the *loxP* vector targeting of an authentic *loxP* site. Furthermore, integration was found to be specific, the presence of the *loxP* site not misdirecting integration of the *mock loxFAS1* plasmid. Importantly, this result demonstrates the feasibility of constructing heterospecific *lox* sites that are proficient for recombination. Also, as shown by Hoess *et al.* (1986) in the prokaryotic system, homology appears to be required for efficient recombination between recombining sites.

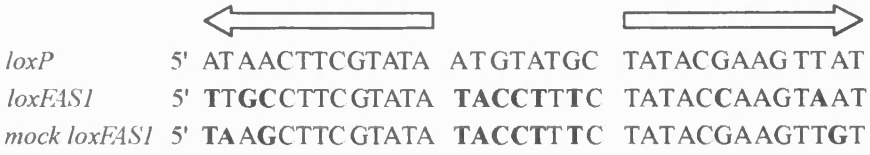


Figure 5.6 Sequences of the *loxP*, *loxFAS1* and modified *mock loxFAS1* site. Differences between the wild-type *loxP* site are highlighted in bold. Arrows show the inverted repeat sequences.

Sauer proposed that by placing multiple, independently acting *lox* sites into the genome at defined locations (by homologous recombination) it might be possible to target different transgene constructs to different chromosomal locations in the same cell, by specifying the corresponding *lox* site on the targeting vector, and that such a strategy would be of particular utility in a mammalian system in which variability of transgene expression is observed depending on the site of transgene integration (the feasibility of using Cre-mediated site-directed integration in a mammalian system had been shown in an earlier study by Fukushige and Sauer [1992]). The intriguing possibility of targeting an endogenous *lox*-like site in the human genome, as an alternative gene therapy strategy, is also suggested. This would eliminate problems due to random DNA integration by permitting precise and stable integration of exogenous DNA at a defined genomic locus.

5.1.2 Experimental strategy

The main drawback of integration via a single cross-over is that plasmid DNA is cointegrated with the desired transgene. A more precise Cre-based method for integrating a foreign DNA into the genome, that avoids the simultaneous introduction of unwanted plasmid DNA, involves a double cross-over event between pairs of different *lox* sites. Recombination results in site-directed exchange of the transgene between the targeting vector and the chromosomal locus (Figure 5.7).

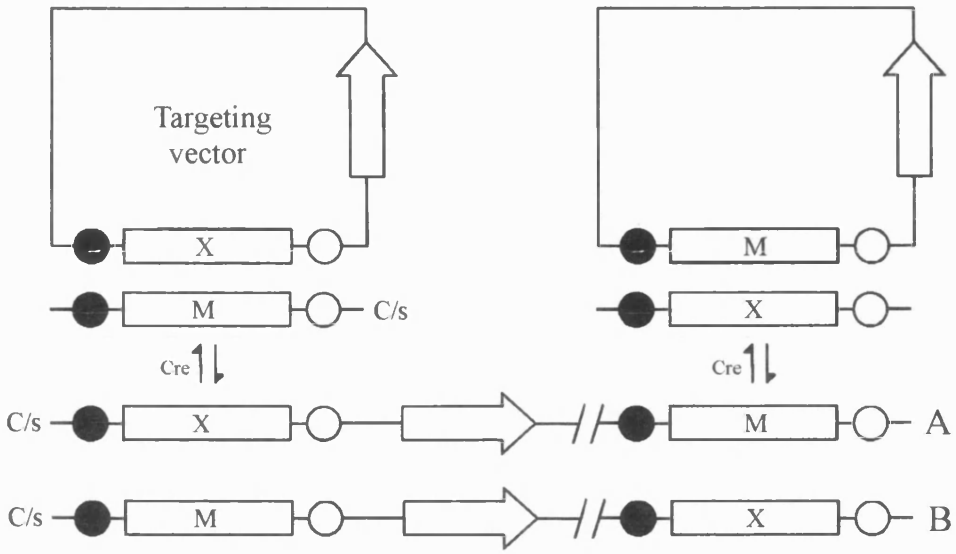


Figure 5.7 Site-specific integration of exogenous DNA via a double cross-over event between two different *lox* sites (circles). M: antibiotic resistance marker. X: DNA segment for insertion. C/s: chromosome. Arrow: nutritional marker gene. The recombination reaction can proceed via two intermediates: **A** and **B**. Intermediate **A** is derived from intermolecular recombination between the filled circle *lox* sites, whilst **B** represents the intermediate formed by crossing-over between the empty circle *lox* sites.

The double recombination strategy has been successfully applied in a mouse cell line (Bethke and Sauer, 1997), but not in yeast. Such a strategy would be valuable for establishing stable (i.e. integrated) DNA repertoires (Figure 5.8). Alternatively it could be used to target several different reporter constructs to a particular locus. The effectiveness of the double recombination system is reliant upon the fidelity of *lox* site recombination, i.e. that pairs of identical *lox* sites recombine efficiently whereas different *lox* sites do not. As discussed, homology has been repeatedly shown to be necessary for efficient recombination. However, Ross-Macdonald *et al.* (1997) reported that the very different *lox* sites, a *loxP* site and a modified *loxR* site (Figure 5.10), *do* undergo efficient (>90%) intramolecular recombination in yeast, suggesting that a Cre-mediated double recombination system might not work in this organism. The aim of the study presented here was to investigate whether a double recombination system could be successfully applied in yeast, and to what extent, if any, undesirable *lox* site recombination might play in reducing the efficiency of the desired reaction.

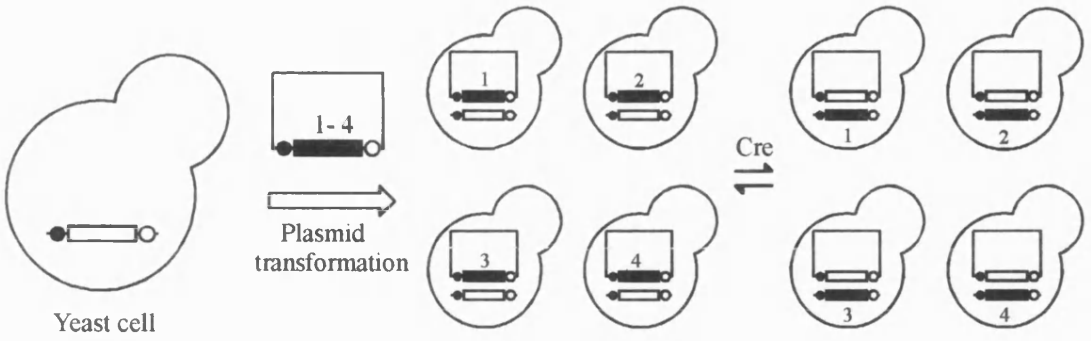


Figure 5.8 Integrating DNA repertoires site-specifically with the double recombination system. *lox* sites are represented by circles. White rectangles: antibiotic resistance marker gene. The numbers refer to different members of the DNA repertoire: black rectangles, which could be a collection of different reporter genes, or a library of mutant genes. Although only four different elements are indicated, the maximum size of the library would in principle only be limited by the efficiency of the plasmid transformation.

Why is it worth developing a double cross-over strategy for yeast, when alternative methods for introducing exogenous DNA elements already exist? The principal current technique, PCR-mediated direct gene replacement (Wach *et al.*, 1996), is summarised in Figure 5.9. In PCR-mediated direct gene replacement and the proposed Cre-mediated double cross-over strategy, the first step is the same: integration of a PCR-generated DNA fragment (containing a counterselectable or selectable marker) at a desired genomic locus. However, a potential limiting step in direct gene replacement is the second transformation reaction to introduce the transgene and to remove the counterselectable marker. Using a lithium acetate protocol and positive selection, Wach *et al.* (1996) reported a transformation efficiency of 138 transformants per μg of PCR-synthesised DNA. In a control transformation with a CEN-ARS plasmid, containing the same positive selection marker, the yield of transformants was 2×10^5 per μg of DNA. It is hoped that the double cross-over strategy, which utilises a CEN-ARS targeting vector, will allow an efficiency of gene replacement - following Cre-mediated exchange - closer to the efficiency obtained for plasmid transformation. This improvement would be particularly useful when introducing DNA libraries into the yeast genome, as shown in Figure 5.8.

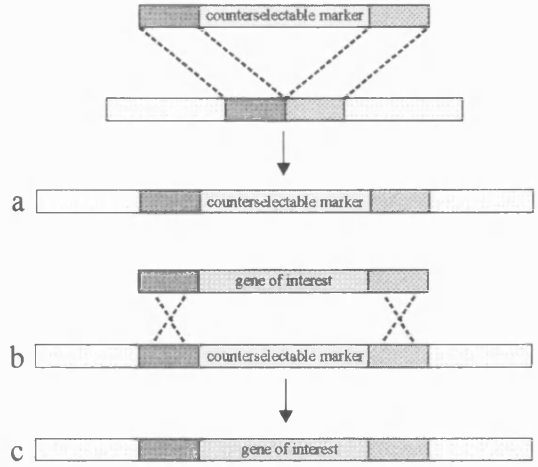


Figure 5.9 PCR mediated direct gene replacement. **a**. The gene to be replaced is first disrupted with a PCR generated DNA fragment containing a counterselectable marker, e.g., *URA3* from *K. lactis*. **b**. The recipient strain is transformed with a second PCR generated fragment containing the transgene. **c**. Counterselection for loss of the marker leads to integration of the transgene. [Note that this method was attempted for the introduction of *OR17-228* at the *HML* locus in the work presented in Chapter 3.]

Naturally the *loxP* and modified *loxR* sites used by Ross-Macdonald *et al.* (1997) would be an unsuitable combination for the double recombination strategy. For this study, the *loxP* site was used in conjunction with one of the mutant *loxP* sites made by Hoess *et al.* (previous section), a site termed *loxP 511* (Figure 5.10). This site, with a base substitution at the +3 position (numbering as Figure 5.2), was shown to be 100% proficient in self × self recombination, but did not show any recombination activity with the *loxP* site in bacteria (Hoess *et al.*, 1986). The *loxP* × *loxP 511* combination was also chosen by Bethke and Sauer for their double cross-over strategy in mouse cells, where no illegitimate recombination between the different *lox* sites was reported.

loxP	5' ATAACTTCGTATA GCATAC AT TATACGAAGT TAT 3'
loxP 511	5' ATAACTTCGTATA **GT**ATAC AT TATACGAAGT TAT 3'
Modified *loxR*	5' ATAACTTCGTATA GCA**GGAG**T TAT**C**CGAAG**C**GAT 3'

Figure 5.10 Sequences of the *loxP*, *loxP 511* and modified *loxR* site. Differences between the wild-type *loxP* site are highlighted in bold. Arrows show the inverted repeat sequences.

The construction of the experimental yeast double cross-over system, based on the model shown in Figure 5.7, is described in detail in the next section (Materials and Methods). The main steps to be described are: **1**. the construction of a gene disruption cassette containing an antibiotic resistance marker (M), flanked by two different *lox* sites (*loxP* and *loxP 511*), and sequences that would allow subsequent integration of the cassette at a chosen chromosomal locus (*LEU2*), by homologous recombination (Wach, 1996). M: the antibiotic resistance marker was the dominant *kanMX4* module (Wach *et al.*, 1994), allowing positive transformants to be selected by their resistance to the antibiotic geneticin (G418); **2**. the construction and transformation of the targeting vector containing the *loxP*, *loxP 511* flanked transgene (X). The transgene was the putative human odorant receptor *OR17-228* (Ben-Arie *et al.*, 1993), chosen in this case as it is heterologous to the yeast genome, eliminating the possibility of misincorporation by general recombination; and **3**. transformation of the experimental strain with a vector capable of expressing Cre, followed by molecular analysis of any recombination events. The vector used contained the *cre* ORF fused to the yeast *GAL1* promoter, allowing Cre production to be regulated by the carbon source. Cre-induced recombination in the model system was monitored by PCR and Southern blot analysis.

5.2 MATERIALS

5.2.1 *Saccharomyces cerevisiae* strains

Supplied *Saccharomyces cerevisiae* strains, and those constructed during the course of the work in this chapter, are described herein.

Strain	Genotype	Reference
FY10	*MATα ura3-52 leu2Δ1 GAL2*$^+$	Winston *et al.*, 1995.
FY22	*MAT***a** *ura3-52 his3Δ200 GAL2*$^+$	Winston *et al.*, 1995.
MRY1	*MAT***a** *his3Δ200 ura3-52 leu2Δ1 lys2Δ202 trp1Δ63 sst2Δ GAL2*$^+$	Redfern, 2000.
CAY1	*MATα sag1Δ::kanMX4 ura3-52 leu2Δ1 GAL2*$^+$	Adams, 1999.
BDY1	*MATα ura3-52 leu2Δ::kanMX4 GAL2*$^+$	This work.
BDY2	*MAT***a** *ura3-52 his3Δ200 leu2Δ::kanMX4 GAL2*$^+$	This work.
BDY3	BDY1::pBS39 *URA3*; pRS315[*OR17-228*] *LEU2*	This work.
BDY4	BDY1::pBS39 *URA3*	This work.
BDY5	*MATα ura3-52 leu2Δ::OR17-228 GAL2*$^+$	This work.

The yeast FY-strains constructed by Winston *et al.* (1995), are a convenient set of *Saccharomyces cerevisiae* that are isogenic to strain S288C (Mortimer and Johnston, 1986), from which all common laboratory strains are derived. The strains are *GAL2*$^+$ and contain non-reverting mutations in genes commonly used for the selection of recombinant plasmids. The strains have been shown to be very strongly induced in the presence of galactose (F. Winston, personal communication). In cells grown in raffinose and induced by the addition of galactose, mRNA levels go from virtually undetectable to greater than 100× increased within 60 minutes. Representative strains from the FY collection were the source of DNA for the European Union Yeast Genome Sequencing Programme.

5.2.2 Plasmids

Supplied DNA vectors used in this chapter are described herein. Plasmids generated during the course of the work in this chapter are described in their relevant sections.

A) *pFA6a-kanMX4*

Plasmid pFA6a-kanMX4 as constructed by Wach *et al.* (1994) is shown in Figure 5.11. The plasmid is one of a series of reporter/marker plasmids (pFA, plasmids for Functional Analysis; Wach *et al.*, 1996). The marker in the vector is the *kanMX4* module which is a hybrid of the coding sequence of the *kan*^{r} gene, from the *E. coli* transposon *Tn903*, and the transcriptional and translational control sequences from the *TEF* gene of the filamentous fungus *Ashbya gossypii*.

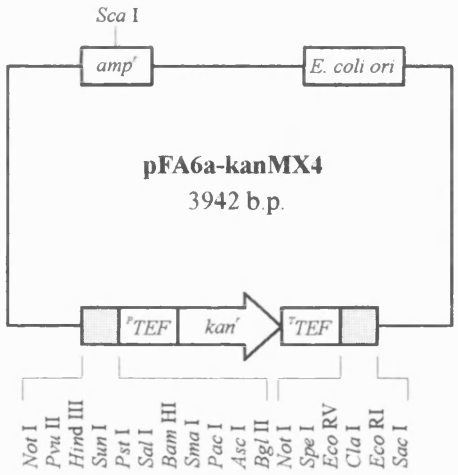


Figure 5.11 Plasmid pFA6a-kanMX4. ^{P}*TEF* and ^{T}*TEF* indicate the promoter and terminator of the *A. gossypii TEF* gene. Shaded boxes represent the multiple cloning site, unique restriction sites are shown (Figure is not to scale).

B) *pRS315-lox D 1.3*

Plasmid pRS315-*lox* D 1.3 was constructed by Adams (1999): A 907 b.p. fragment amplified from plasmid pYX243-*lox* D 1.3 (not shown) was inserted (after removal of a *URA3* marker) between the *Apa*I and *Not*I sites of vector pTG5756 (Section 3.2.2c) to yield pRS315-*lox* D1.3 (Figure 5.12).

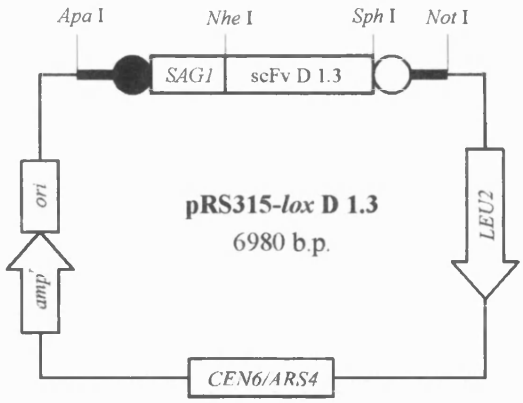


Figure 5.12 Plasmid pRS315-*lox* D 1.3. Closed and open circles represent *lox P* and *lox P 511* sites respectively. Thick lines represent pYX243-*lox* D 1.3 vector sequence. *SAG1* and scFV D 1.3 are as described in the main text. Relevant unique restriction sites are shown (Figure is not to scale).

The cloned fragment contains the leader sequence (first 57 nucleotides from start of the initiation codon) of the *S. cerevisiae SAG1* gene, spliced to a single chain antibody fragment D 1.3 (scFv D 1.3). The gene fusion is flanked on one side by a *loxP* site and on the other by a *loxP 511* site. Also present at either end of the cloned fragment are remnants of the parent vector pYX243-*lox* D 1.3. Relevant sequence information about the cloned fragment, including orientation of the *lox* sites is presented in Figure 5.13.

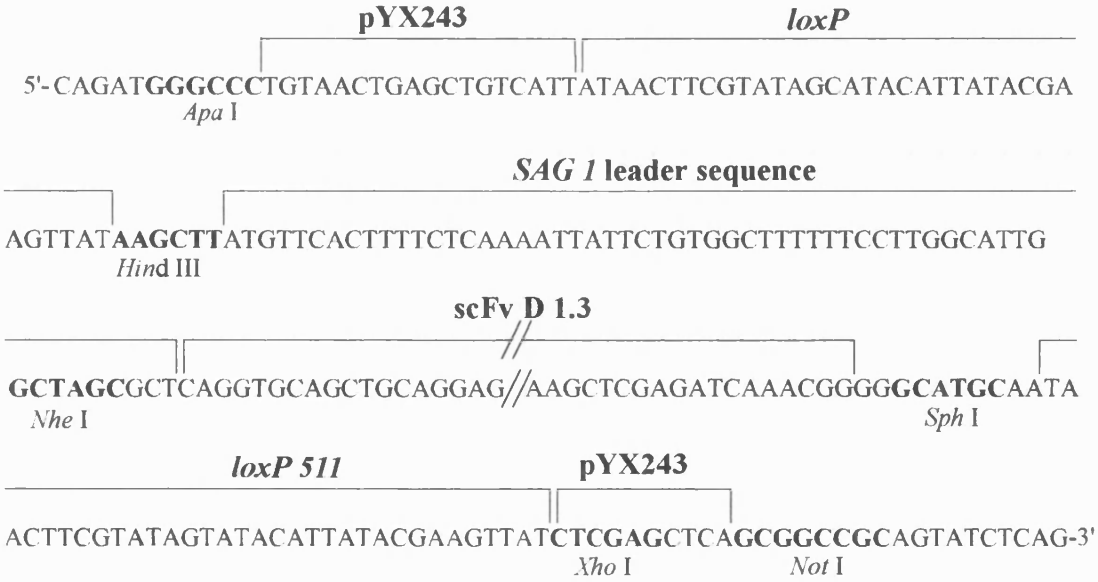


Figure 5.13 Sequence of *lox* flanked *SAG1*-leader/scFV D 1.3 gene fusion cassette. Relevant restriction sites to this work are represented in bold. The *Apa*I and *Not*I sites at the 5' and 3' ends of the fragment respectively were those used by Adams (1999) for cloning the fragment into pRS315-*lox* D 1.3.

C) *pBS39*

Plasmid pBS39 as constructed by Sauer (1987), is shown in Figure 5.14. The vector contains the *cre* gene of coliphage P1 under the control of the yeast *GAL1* promoter region. The *GAL1* gene is expressed at high levels in galactose but is completely repressed in the presence of glucose, allowing controlled expression of *cre* recombinase.

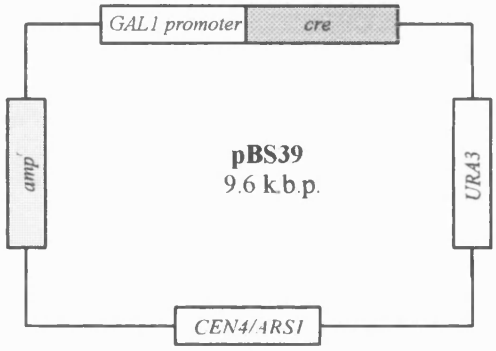


Figure 5.14 The Cre recombinase expression vector pBS39 (Figure is not to scale).

5.3 METHODS: GENERATION OF EXPERIMENTAL STRAINS FOR THE ANALYSIS OF CRE-MEDIATED RECOMBINATION

5.3.1 Construction of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette

The gene disruption cassette was constructed in two stages. Firstly the *kanMX4* module from vector pFA6a-kanMX4 was sub-cloned into vector pRS315-*lox* D1.3 in order to flank the marker module with the required *lox* sites. In the second step the *loxP-kanMX4-loxP 511* fragment was excised from the new vector and joined via PCR to the *LEU2* flanking sequences, generating the completed cassette. Both stages are described below.

Stage 1: Construction of the vector pRS315-loxP-kanMX4-loxP 511

A) *Primers to amplify the kanMX4 module from pFA6a-kanMX4*

The sequences of both primers (Table 5.1) used to amplify the *kanMX4* module were derived from the multiple cloning site of vector pFA6a-kanMX4 (Section 5.2.2a). The DNA sequences were obtained from Wach *et al.* (1994; Figure 1).

The fundamental design parameters for primers were as described in Section 2.2.1. The primers kanMX4F(Nhe) and kanMX4F(Sph) were designed to anneal to the pFA6a-kanMX4 vector either side of the *kanMX4* module, encompassing the fragment of the plasmid recommended for the replacement of target open reading frames (Wach *et al.*, 1996).

Both primers contained either a *Not*I or *Sph*I restriction endonuclease recognition site at their 5' ends to allow subsequent cloning of the amplified fragment into vector pRS315-lox D 1.3 (Section 5.2.2b).

B) *PCR amplification of the kanMX4 module*

An Expand™ PCR (Section 2.2.3b) using the primers kanMX4F(Nhe) and kanMX4(Sph) was used to amplify a 1541 b.p. fragment from vector pFA6a-kanMX4. The reaction was performed for 25 cycles using the following parameters: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min]. The PCR product was quantified by agarose gel electrophoresis (Section 2.2.4).

Primer name	Sequence (5' – 3')	Description (5' – 3')
kanMX4F(Nhe)	GCGATGAACT**GCTAGC**G TACGCTGCAGGTCGACG	10 n.t. overhang; a *Nhe*I site; 18 n.t. homology to vector pFA6a-kanMX4. Binds to template strand of *kanMX4* cassette, 433 to 416 b.p. upstream from *kan*^{r} gene initiation codon.
kanMX4F(Sph)	GCGATGAACT**GCATGC**T ATCATCGATGAATTCGA	10 n.t. overhang; a *Sph*I site; 18 n.t. homology to vector pFA6a-kanMX4. Binds to coding strand of *kanMX4* cassette, 249 to 266 b.p. downstream from *kan*^{r} gene termination codon.

Table 5.1 Sequences of the oligonucleotides used to amplify the *kanMX4* cassette from vector *pFA6a-kanMX4*. Restriction endonuclease sites are in bold.

C) *Cloning of the kanMX4 module into pRS315-lox D 1.3*

Plasmid pRS315-*lox* D 1.3 (~1 µg) [Section 5.2.2b] was double-digested (Section 2.2.14) with *Nhe*I (5 units) and *Sph*I (5 units) at 37°C overnight. The vector was purified from the excised fragment (the single chain antibody fragment D 1.3) by low melting point agarose gel electrophoresis (Section 2.2.4a) and recovered by β-agarase digestion (Section 2.2.4b). The amplified *kanMX4* module (from above) was digested and purified in the same manner. Both double-digested products were quantified by agarose gel electrophoresis (Section 2.2.4). The *kanMX4* module (~20 ng) was then ligated into the vector (~30 ng)[Section 2.2.15] and transformed directly by electroporation into competent *E. coli* (Section 2.2.9d). Ampicillin resistant colonies were PCR screened (Section 2.2.12a) for the presence of the *kanMX4* module using primers T7 and M13 (Table 5.2). PCR products were analysed by agarose gel electrophoresis (Section 2.2.4). Positive transformants were grown overnight in 4 ml of LB (Section 2.2.6) from which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

The construct from one transformant was checked for correct integration and for the presence of any mutations by sequencing (Section 2.2.16) using primers T7 and M13.

Primer name	Sequence (5' - 3')	Description (5' – 3')
T7	GTAATACGACTCACTATAGGGC	Binds to template strand of pRS-vector, relative co-ordinates on coding strand, 3096 - 3117 b.p. (Sikorski and Hieter, 1989).
M13	GGAAACAGCTATGACCATG	Binds to coding strand of pRS-vector, co-ordinates 3297 - 3278 b.p. (Sikorski and Hieter, 1989).

Table 5.2 Sequences of the oligonucleotides used for screening and sequencing of vector pRS315-*loxP-kanMX4-loxP 511*.

Stage 2: Generating the completed gene disruption cassette

D) *Construction of the loxP-kanMX4-loxP 511 cassette*

The vector pRS315-*loxP-kanMX4-loxP 511* (~0.1 µg) was digested (Section 2.2.14) overnight with *Apa*I (10 units) at 37°C. The cut vector was purified using the appropriate DNA extraction kit (Section 2.2.4c) and then further digested with *Not*I (10 units). The released *loxP-kanMX4-loxP 511* cassette was purified using low melting point agarose gel electrophoresis (Section 2.2.4a). The cohesive ends of the cassette generated by the double restriction digest were end-filled using Expand™ (dNTPs and 10× Expand™ buffer were added to the DNA to the same concentration as for an Expand™ PCR [Section 2.2.3b] and the reaction incubated at 72°C in the thermal cycler for 5 min.). The resultant end-filled fragment was quantified by agarose gel electrophoresis (Section 2.2.4).

E) *Primers to amplify LEU2 flanking sequences*

The sequences of the four primers (Table 5.3) used to generate the disruption cassette were derived from the promoter and terminator regions of the *LEU2* locus. The DNA sequences were obtained from the *Saccharomyces cerevisiae* genome database (Section 2.2.1). The accession number for the *LEU2* open reading frame is YCL018W.

The Leukan(F) and Leukan(B) primers were designed to have their 5' ends homologous to 24 nucleotides at each end of the *loxP-kanMX4-loxP 511* cassette. This was to allow the *LEU2* flanking sequences, once amplified, to be joined to the central portion of cassette via a 'splicing by overlap extension' PCR. The scheme to generate the completed disruption cassette is illustrated in Figure 5.15.

Primer name	**Sequence (5' – 3')**	**Description (5' – 3')**
Leu2PF2	ACAGGGGCGCTATCGCACAG	Binds to the template strand, 321 to 302 b.p. upstream from *LEU2* initiation codon.
Leukan(F)	GAAGTTATCTCGAGCTCAGCG GCCCTGTCGCCGAAGAAGTTA AG	24 n.t. homology to end-filled *Not*I end of *loxP-kanMX4-loxP 511* cassette; binds to template strand with respect to *kan*^{r} gene. 20 n.t. homology to template strand, 35 to 15 b.p. upstream from *LEU2* termination codon.
Leu2TB2	GAGAAAATCCTCCAATATAT	Binds to the coding strand, 325 to 305 b.p. downstream from *LEU2* termination codon.
Leukan(B)	GTTATAATGACAGCTCAGTTA CAGTAGAATGGTATATCCTTG AA	24 n.t. homology to blunt-ended *Apa*I end of *loxP-kanMX4-loxP 511* cassette; binds to coding strand with respect to *kan*^{r} gene. 20 n.t. homology to coding strand immediately before *LEU2* initiation codon.

Table 5.3 Sequences of the oligonucleotides used in the generation of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette.

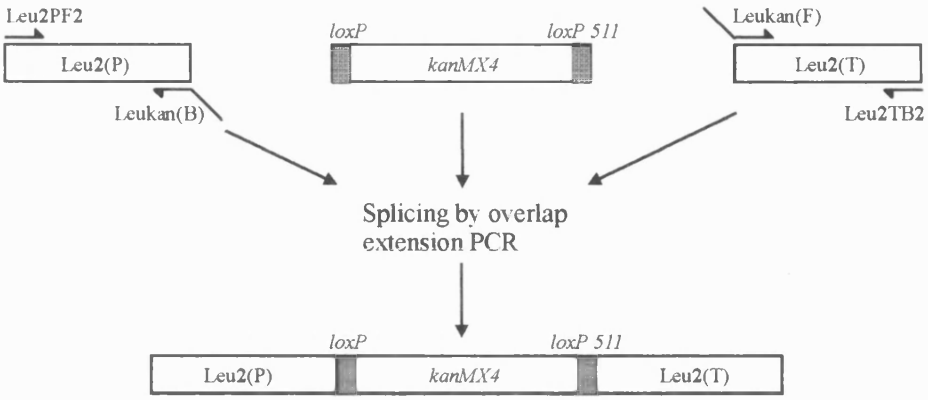


Figure 5.15 Schematic representation of the synthesis of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette. Leu2(P): the promoter region of the *LEU2* gene. Leu2(T): the terminator region of the *LEU2* gene (note that the terminator region contains the last 35 b.p. of the *LEU2* ORF). The amplified promoter and terminator regions were joined to the *lox* flanked *kanMX4* cassette (excised from vector pRS315-*loxP-kanMX4-loxP 511*) via a splicing by overlap extension PCR.

F) *PCR amplification of the LEU2 flanking sequences*

Two separate Expand™ PCRs (Section 2.2.3b) were performed to amplify the promoter (P) and terminator (T) regions of the *LEU2* gene from yeast genomic DNA. The genomic DNA was prepared from the *S. cerevisiae* **a**-strain MRY1 (Section 5.2.1) following the procedure outlined in Section 2.2.13.

The primer combination Leu2PF2/Leukan(B) was used to amplify the promoter region of the *LEU2* gene. The reaction was performed for 25 cycles using the following parameters: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min].

The terminator region was amplified using the primer combination Leukan(F)/Leu2TB2 with the same PCR condition as above.

Both PCR products were purified by LMP-agarose gel electrophoresis (Section 2.2.4a) and recovered by digestion with β-agarase I (Section 2.2.4b). The purified fragments were quantified by agarose gel electrophoresis (Section 2.2.4).

G) *PCR mediated generation of the LEU2 disruption cassette*

Approximately equal amounts (~50 ng) of the *LEU2*-promoter fragment, the *LEU2*-terminator fragment and the end-filled *loxP-kanMX4-loxP 511* cassette were added together in an Expand™ PCR (Section 2.2.3b) with the primer pair

Leu2PF2/Leu2TB2. The reaction conditions were: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 2 min) × 10; (94°C, 15 s; 55°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min]. The assembled PCR product (2.3 k.b.p.) was separated from the other non-specific bands by LMP-agarose gel electrophoresis (Section 2.2.4a) and recovered by digestion with β-agarase I (Section 2.2.4b).

5.3.2 Integration of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette into *Saccharomyces cerevisiae* strains FY10 and FY22

A) *Transformation of the gene disruption cassette into FY10 and FY22*

The assembled *LEU2* disruption cassette (previous section) was transformed into two yeast strains, the **α**-strain, FY10 and the **a**-strain, FY22 (Section 5.2.1) using a lithium acetate/polyethyleneglycol transformation protocol (Section 2.2.10b). Approximately 0.1 μg of DNA was used to transform each strain. Transformants were inoculated onto YPD plates (Section 2.2.7) containing G418 (400 μg/ml) and incubated at 30°C for 3 days.

B) *Verification of correct cassette integration by PCR*

Two geneticin resistant transformants from each strain were selected from which glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Both novel joints (Figure 5.16) formed upon integration were screened via *Taq* PCRs (Section 2.2.3a) using two sets of primers (Table 5.4).

Correct integration of the *LEU2*-promoter end of the disruption cassette into the genome was checked using the primer pair Leu2PF3/FL2B. The cycling parameters were: [94°C, 2 min; (94°C, 30 s; 50°C, 1 min; 72°C, 1 min) × 30].

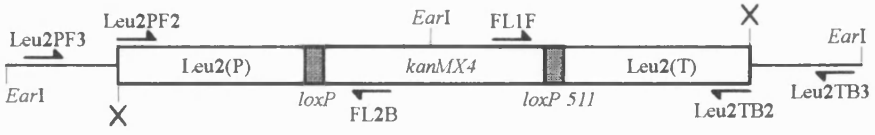


Figure 5.16 Diagrammatic representation showing the position and orientation of the primers used to verify correct cassette integration at the *LEU2* locus. Symbol '×' represents the junction between the introduced DNA fragment and the chromosomal DNA (thin line)[Figure is not to scale].

Integration events at the *LEU2*-terminator end of the cassette were examined using the primers FL1F and Leu2TB3. Reaction conditions were the same as above. All PCR products were analysed by agarose gel electrophoresis (Section 2.2.4).

Primer name	**Sequence (5' – 3')**	**Description (5' – 3')**
Leu2PF3	GAATACTCAGGTATCGTAAG	Binds to the template strand, 400 to 381 b.p. upstream from the *LEU2* initiation codon.
FL2B	GGCGACAGTCACATCATGCC	Binds to the coding strand of the *kanMX4* cassette, 272 to 291 b.p. upstream from the initiation codon.
Leu2TB3	TGAGGGAACTTTCACCATTA	Binds to the coding strand, 435 to 415 b.p. downstream from the *LEU2* termination codon.
FL1F	CCTCGACATCATCTGCCC	Binds to the template strand of the *kanMX4* cassette, 115 to 133 b.p. downstream from the termination codon.

Table 5.4 Sequences of the oligonucleotides used to verify correct cassette integration at the *LEU2* locus.

C) *Verification of correct cassette integration by Southern blotting*

Southern blot analysis was performed on one of the transformants from each new strain (above Section). The oligonucleotide FL2B (Table 5.4) was chosen as the probe as it binds specifically to the *loxP-kanMX4-loxP 511* cassette. The restriction endonuclease *Ear*I was chosen to digest the genomic DNA as it cuts the target sequence to produce a fragment (1339 b.p.) that should be resolved well by the Southern blot. *Ear*I cuts just outside (~200 b.p.) either end of the introduced DNA

fragment and once internally, 238 b.p. downstream from the *kan*^{r} gene's initiation codon. The relative positions of the *Ear*I cut sites (and the FL2B binding site) are shown diagrammatically in Figure 5.16.

The agarose gel was prepared as described in Section 2.2.20a. A comb was used producing wells with approximate dimensions of 3 × 13 × 10 mm.

Fresh genomic DNA was prepared of each new strain and from wild-type FY10 and FY22 (to be used as negative controls for the analysis). The DNA was prepared as described in Section 2.2.13, except after the final resuspension step undissolved material was removed by centrifugation (13 k.r.p.m. for 5 min.).

*Ear*I (3µl, 30 units), 10× NEB reaction buffer 1 (10 µl)[10 mM Bis Tris Propane HCl, 10 mM $MgCl_2$, 1mM dithiothreitol (pH 7.0 @ 25°C)] and water (37µl) were gently mixed with the genomic DNA preparations (50 µl). The digests were incubated overnight at 37°C.

6× LMP-agarose gel loading buffer (20 µl) was mixed with each sample and the samples loaded onto the gel. The gel was run for 12 hours at 30 volts (~1 V/cm).

The DNA was transferred to the nylon membrane as described in Section 2.2.20b.

FL2B, was end-labelled with ^{32}P using T4 polynucleotide kinase (PNK) and [γ-^{32}P]ATP: FL2B (30 pmol), PNK (1 µl, 10 Richardson units/µl), [γ-^{32}P]ATP (9 µl, 90 µCi), 10× PNK buffer (6 µl)[70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol] were made up to a final volume of 60 µl with water. The reaction was incubated at 37°C for 1 hour (adapted from Tomlinson *et al.*, 1992). To remove unincorporated [γ-^{32}P]ATP the reaction was alcohol precipitated (Section 2.2.5) and resuspended with water (60 µl). Percentage incorporation of ^{32}P into FL2B was determined by performing a scintillation count: 1 µl was removed before and after alcohol precipitation, added to scintillation fluid (10 ml) and counted (at 40%) on a NE-6500 liquid scintillation counter.

Pre-hybridisation, hybridisation and film development (exposed for 3 days) were as described in Section 2.2.20c

D) *Verification of the disruption cassette in FY10 by DNA sequencing*

The primer pair Leu2PF3/Leu2TB3 was used to amplify the locus containing the integrated disruption cassette from FY10 genomic DNA. An Expand™ PCR

(Section 2.2.3b) with the following parameters was performed: [94°C, 2 min; (94°C, 15 s; 50°C, 30 s; 72°C, 2min) × 10; (94°C, 15 s; 50°C, 30 s; 72°C, 2 min + 20 s/cycle) × 15; 72°C, 5 min]. The PCR product was analysed by agarose gel electrophoresis (Section 2.2.4). The sample was then purified by LMP-agarose gel electrophoresis (Section 2.2.4a) and extracted by β-agarase digestion (Section 2.2.4b).

Two separate DNA sequencing reactions (Section 2.2.16) were carried out to check the integrity of the *lox* sites in the integrated cassette. The primer Leu2PF2 was used to analyse the *loxP* site and the primer Leu2TB2 to analyse the *loxP 511* site.

5.3.3 Construction of targeting vector pRS315[*OR17-228*]

A) *Primers to amplify OR17-228*

The primers Leu2XbaI and OR17-228(Sph) were used to amplify the open reading frame of the putative human odorant receptor, *OR17-228*, from a previously made DNA fragment, the *HML* disruption cassette (Section 3.3.6d). The sequences of the primers (Table 5.5) were derived from the start of the *OR17-228* ORF and the terminator region of the *LEU2* gene. The *HML* disruption cassette is shown again, with the position of the primers, in Figure 5.17. The DNA sequences for the primers were obtained from the SGD (for Leu2XbaI) and Ben-Arie *et al.* (1993) [for OR17-228(Sph)]. For subsequent cloning steps each primer had either a *Sph*I or *Xba*I restriction endonuclease recognition site in its 5' end.

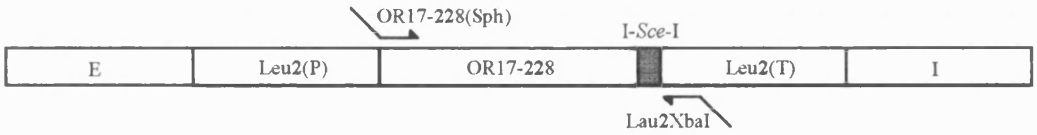


Figure 5.17 The *HML* locus gene disruption cassette showing the position of the primers used to amplify the *OR17-228* ORF for cloning into the targeting vector pRS315 (Figure is not to scale).

B) *Amplification of OR17-228*

Using the OR17-228(Sph)/Leu2XbaI primer combination, an Expand™ PCR (Section 2.2.3b) was performed to amplify the *OR17-228* ORF. The DNA was the *HML* disruption cassette. The reaction was performed for 25 cycles using the following parameters: [94°C, 2 min; (94°C, 15 s; 55°C, 30 s; 72°C, 1 min) × 10;

(94°C, 15 s; 55°C, 30 s; 72°C, 1 min + 20 s/cycle) × 15; 72°C, 5 min]. The PCR product was analysed (Section 2.2.4) and purified by LMP-agarose gel electrophoresis (Section 2.2.4b).

C) *Cloning of OR17-228 into pRS315-lox D 1.3*

The purified PCR product (~1.2 µg) was double-digested (Section 2.2.14) in a total digest volume of 100 µl with *Xba*I (40 units) and *Sph*I (20 units) at 37°C overnight; then purified by alcohol precipitation (Section 2.2.5) and quantified (Section 2.2.4). The plasmid pRS315-*lox* D1.3 (Section 5.2.2b) was prepared as described in Section 5.3.1c. The *OR17-228* insert (~180 ng) was then ligated into the vector (~50 ng)[Section 2.2.15] and 2 µl transformed directly into competent *E. coli* TOP10F' cells (Section 2.2.9d). Ampicillin resistant colonies were PCR screened (Section 2.2.12a) using the T7/M13 (Table 5.2) primer combination. PCR products were analysed by agarose gel electrophoresis (Section 2.2.4). Positive transformants were grown overnight in 4 ml of LB (Section 2.2.6) from which glycerol stocks (Section 2.2.8) and plasmid DNA (Section 2.2.17) were prepared.

The construct from one transformant was checked for correct integration and for the presence of any mutations by sequencing (Section 2.2.16) using the primers, T7 and M13.

Primer name	**Sequence (5' - 3')**	**Description (5' - 3')**
OR17-228(Sph)	CTGAGATACT**GCATGC**AT GGAGCCAGAAGCTGGG	10 n.t. overhang; a *Sph*I site; 18 n.t. homology to coding strand of *OR17-228* ORF starting at the initiation codon
Leu2XbaI	GCTACC**TCTAGA**TTAACTT CTTCGGCGACA	6 n.t. overhang; a *Xba*I site; 18 n.t. homology to template strand of *LEU2* locus, 15 - 35 b.p. upstream from *LEU2* termination codon.

Table 5.5 Sequences of the oligonucleotide primers used to amplify the *OR17-228* ORF from the *HML* disruption cassette. Restriction endonuclease recognition sites are in bold.

5.3.4 Transformation of the *LEU2* disrupted FY10 strain with pBS39 and pRS315[*OR17-228*]

The vector pRS315[*OR17228*] (~100 ng) and the Cre recombinase expression vector pBS39 (~100 ng)[Section 5.2.2c], were co-transformed (Section 2.2.10d) into the FY10 strain containing the *loxP-kanMX4-loxP 511* cassette integrated at the *LEU2* locus (BDY1; Section 5.2.1). Ura^+ Leu^+ transformants were selected and then grown at 30°C for 24 hours in SC-Ura-Leu (10 ml), from which glycerol stocks (Section 2.2.8) and genomic DNA were prepared (Section 2.2.13). Vector pBS39 was also transformed individually into BDY1. Ura^+ colonies were selected and grown at 30°C for 24 hours in SC-Ura (10 ml), and again glycerol stocks and genomic DNA were prepared from these cultures.

Genomic DNA from the transformants was screened via *Taq* PCRs (Section 2.2.3a); plasmid pBS39 was detected with the primers pYX(FOR2) and Cre(Back) [Table 5.6] and pRS315[*OR17-228*] was detected using primers T7 and M13 [Table 5.2]. For both reactions the cycling parameters were: [90°C, 2 min; (90°C, 30 s; 50°C, 30 s; 72°C, 1 min) × 30]. All PCR products were analysed by agarose gel electrophoresis (Section 2.2.4).

Primer name	Sequence (5' - 3')	Description (5' – 3')
Cre(Back)	TTGAAGGCTCTCAAGGGC	Binds to the coding strand, 4 b.p. after the *Sal*I site on vector pBS39.
pYX(FOR2)	TGTATTACTTCTTATTCA	Binds to the template strand in the *GAL1* promoter region of vector pBS39.

Table 5.6 Sequence of the oligonucleotide primers used to screen for the presence of the Cre recombinase expression vector pBS39.

5.4 METHODS: CRE-MEDIATED RECOMBINATION

5.4.1 Liquid galactose induction of Cre recombinase

A single colony of BDY3 (Section 5.2.1) was picked from an SC-Leu-Ura plate (2% glucose) into 1 ml of the same medium and grown for 24 hours at 30°C, with shaking at 800 r.p.m. The cells were washed twice with 1 ml of water and resuspended in 1 ml of water. Cre recombinase production was then induced by inoculating 10 ml of fresh liquid SC-Leu-Ura (2% galactose) medium with 5×10^7 resuspended cells and incubating for 24 hours at 30°C, with shaking at 250 r.p.m. After this induction period, approximately 300 cells were spread on an SC-Leu plate which was then incubated for 3 days at 30°C. A control experiment in which 5×10^7 resuspended cells were transferred to fresh liquid SC-Leu-Ura (2% glucose) medium instead of SC-Leu-Ura (2% galactose), with subsequent growth of cells on an SC-Leu plate, was also performed. One hundred colonies from each SC-Leu plate were picked and streaked onto a YPD plate containing 200 μg/ml G418.

5.4.2 Analysis of recombination events by PCR screening

A) *Screening primers used*

PCR screens of the 100 Leu^+ colonies were mostly carried out using the primers LEU2PF3 and LEU2TB3 (Table 5.4), which were designed to amplify the entire *leu2* locus. These primers also amplified the *LEU2* gene from the pRS315[*OR17-228*] plasmid (see Discussion). Plasmid screens were carried out with primers T7 and M13 (Table 5.2).

The locations of the oligonucleotide binding sites with respect to the model system are shown in Figure 5.18.

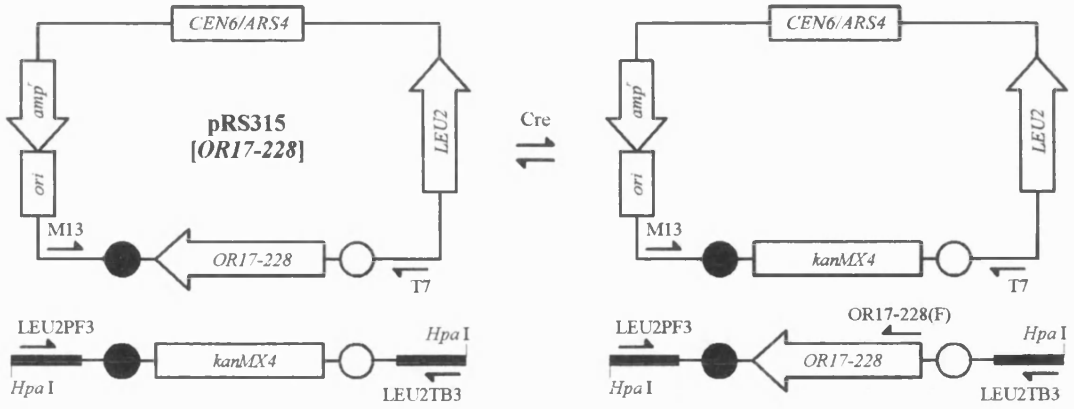


Figure 5.18 The vector and target structures of the Cre-mediate double recombination system showing positions of the oligonucleotides and *Hpa*I cut sites used for PCR screening and Southern blot analysis. Closed and open circles represent *loxP* and *loxP 511*, sites respectively. Thick lines represent part of the *leu2* locus (Figure is not to scale).

B) *Amplification of target sequences*

The whole-cell PCR screening method (Section 2.2.12b) was used to amplify the *leu2* locus and vector sequence from each of the Leu$^+$ colonies. The *leu2* locus was amplified using the primer pair LEU2PF3/LEU2TB3 with the following PCR cycling parameters: [94°C, 2 min; (94°C, 30 s; 50°C, 30 s; 72°C, 2 min 30 s) × 30]. The vector sequence was amplified using the primer pair T7/M13 with cycling parameters as above, except the annealing temperature was 48°C.

C) *PCR analysis of recombination intermediates*

The primers GCvCO and LEU2TB4 (Table 5.7) were used to determine whether intermediates in the exchange process were due to integration of the pRS-targeting vector at the *leu2* locus. LEU2TB4 was designed with reference to the terminator region of the *leu2* locus, crucially binding to the *LEU2* sequence not present in the pRS-targeting vector. GCvCO was designed with reference to the pRS-targeting vector, binding immediately before the start of the *LEU2* sequence. The positions of the primer binding sites is shown diagrammatically in Figure 5.25.

Genomic DNA (Section 2.2.13) from the colonies of interest was prepared and a *Taq* PCR (Section 2.2.3a) performed. The cycling parameters for the reaction were: [94°C, 2 min; (94°C, 30 s; 48°C, 30 s; 72°C, 3 min) × 30]. All PCR products were analysed by agarose gel electrophoresis (Section 2.2.4).

Primer name	Sequence (5' - 3')	Description (5' - 3')
GCvCO	TCTGTGCGGTATTTCACACC	Binds to coding strand of pRS-vector, co-ordinates 2467 - 2487 b.p. (Sikorski and Hieter, 1989).
LEU2TB4	GGTAAGGATGATGCATTAGC	Binds to the coding strand, 547 to 527 b.p. downstream from the *LEU2* termination codon.

Table 5.7 Oligonucleotides used to analyse intermediates in Cre-induced exchange.

5.4.3 Analysis of recombination events by DNA sequencing

The method for sequencing the *lox* sites at the *leu2* locus in Cre-induced strains is the same as that described for verifying the disruption cassette in FY10 (Section 5.3.2d).

For sequencing of the *lox* site in colonies where the DNA element had been deleted, the template was amplified as above, except that the extension time was reduced to 1 minute. The primer LEU2PF2 was then used in the sequencing reaction.

5.4.4 Southern blot analysis of exchangeants

Southern blot analysis was performed on 5 colonies that PCR screening and DNA sequencing revealed had undergone the desired exchange event. The oligonucleotide OR17-228(F) (Table 3.5) was chosen as the probe as it binds specifically to the integrated DNA element; the *OR17-228* ORF. The restriction endonuclease *Hpa*I was chosen to digest the genomic DNA; this enzyme makes a double-stranded cut either side of the integrated *lox*-flanked *OR17-228* ORF (and outside the *LEU2* segment employed for integration of the *kanMX4* module), to produce a fragment (2281 b.p.) that should be resolved well by the Southern blot gel. The positions of the OR17-228(F) binding site and the *Hpa*I cut sites is shown diagrammatically in Figure 5.18.

The agarose gel was prepared as described in Section 2.2.20a. A comb was used producing wells with approximate dimensions of 3 × 5 × 10 mm.

Before preparing the genomic DNA for analysis the pRS315 vector was outgrown in the 5 colonies of interest (see Discussion). Vector outgrowth was achieved by successively growing the colonies in non-selective medium [each colony was picked into YPD (10 ml) and grown overnight (30°C, 250 r.p.m.), from the resulting culture 10 µl was transferred to fresh medium and grown as before. This was repeated 3 times before selecting for 'vector-loss' colonies by plating onto YPD medium, from which colonies were picked onto SC-Leu medium, in order to select non-growing (vector-loss) colonies; designated as BDY5]. Two genomic DNA samples were prepared for each of the five colonies (Section 2.2.13), these were combined into a volume of 40 µl using the Promega Wizard™ DNA Clean-Up Kit (Section 2.2.4c) [this step was done as it had originally been intended to digest the genomic DNA in as small a volume as possible and to load the sample directly onto the Southern blot gel (see Discussion): instead a larger digest volume (see below) was used and the DNA resuspended in a smaller volume by alcohol precipitation.

*Hpa*I (5µl, 25 units), 10× NEB reaction buffer 4 (10 µl)[50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate, 1mM dithiothreitol (pH 7.9 at 25°C)] and water (45 µl) were gently mixed with the genomic DNA preparations (40 µl). The digests were incubated overnight at 37°C.

The *Hpa*I digested DNA was alcohol precipitated (Section 2.2.5) and re-suspended in water (42 µl), 6× LMP-agarose gel loading buffer (8 µl) was mixed with each sample and the samples loaded onto the gel. The gel was run for 12 hours at 30 volts (~1 V/cm).

The DNA was transferred to the nylon membrane as described in Section 2.2.20b.

OR17-228(F) was end-labelled with ^{32}P using T4 polynucleotide kinase (PNK) and [γ-^{32}P]ATP: OR17-228(F) [15 pmol], PNK (1 µl, 10 Richardson units/µl), [γ-^{32}P]ATP (3 µl), 10× PNK buffer (3 µl)[70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol] were made up to a final volume of 30 µl with water. The reaction was incubated at 37°C for 30 minutes (adapted from Tomlinson *et al.*, 1992). The end-labelled probe was then added directly to the hybridisation buffer (see Discussion).

Pre-hybridisation, hybridisation and film development (exposed for 3 days) were as described in Section 2.2.20c.

5.5 RESULTS: GENERATION OF EXPERIMENTAL STRAINS FOR THE ANALYSIS OF CRE-MEDIATED RECOMBINATION

5.5.1 Construction of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette

Stage1: construction of vector pRS315-loxP-kanMX4-loxP 511

Vector pRS315-*loxP-kanMX4-loxP 511* was constructed by sub-cloning the *kanMX4* module from vector pFA6a-kanMX4 (Section 5.2.2a) into vector pRS315-*lox* D 1.3 (Section 5.2.2b).

The *kanMX4* module was amplified from pFA6a-kanMX4 (Section 5.3.1b). The size of the generated PCR fragment was determined by agarose gel electrophoresis and was in accordance with the predicted theoretical size of 1541 b.p.

Vector pRS315-*lox* D 1.3 was digested with *Nhe*I and *Sph*I in order to remove the single chain antibody fragment scFv D 1.3 (Section 5.3.1c). The exact position of the restriction sites used is shown in Figure 5.2. The *kanMX4* module with *Nhe*I/*Sph*I cohesive ends was then ligated into the vector and transformed directly into *E. coli.* Five ampicillin resistant colonies were PCR screened and two produced the expected 1708 b.p. band, corresponding to the sub-cloned *kanMX4* module. The construct from one of these colonies was sequenced and shown to contain a correctly integrated *kanMX4* cassette with no mutation. The constructed vector is as for pRS315-*lox* D 1.3 (Figure 5.12) except that the single chain antibody fragment (scFv D 1.3) has been replaced by the *kanMX4* cassette.

Stage 2: Generating the completed gene disruption cassette

The gene disruption cassette was constructed by means of a 'splicing by overlap extension' reaction (Figure 5.15) using the *loxP-kanMX4-loxP 511* cassette (excised from plasmid pRS315-*loxP-kanMX4-loxP 511*) in conjunction with two sequences amplified from the *LEU2* locus.

The *loxP-kanMX4-loxP 511* cassette was released from vector pRS315-*loxP-kanMX4-loxP 511* by sequential restriction digestion with *Apa*I and *Not*I (Section 5.3.1d). The cohesive ends of the cassette were filled in, and the size of the resultant fragment checked by agarose gel electrophoresis. The size of the fragment was in accordance with that expected, 1679 b.p.

The two regions of the *LEU2* locus to be used as flanking sequences in the disruption cassette were individually amplified from yeast genomic DNA (Section 5.3.1f). The sizes of both the generated PCR fragments were determined by agarose gel electrophoresis and were in agreement with the predicted theoretical sizes, i.e. 345 b.p. for the promoter region and 384 b.p. for the terminator region.

The three fragments were then joined to one another using a 'splicing by overlap extension' reaction. This was possible due to the presence of the 24 nucleotides of homology each of the flanking DNA fragments contained to the *loxP-kanMX4-loxP 511* cassette. The completed *LEU2* gene disruption cassette was obtained whose size was in agreement with that expected, 2360 b.p.

5.5.2 Integration of the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette into *Saccharomyces cerevisiae* strains FY10 and FY22

The gene disruption cassette (previous section) was transformed into two yeast strains, FY10 and FY22, using a lithium acetate/polyethyleneglycol transformation procedure (Section 5.3.2a). After three days incubation numerous transformants (~500 colonies/plate) were observed growing for both strains, corresponding to approximately 10 geneticin resistant ($G418^R$) transformants per ng of DNA.

Two $G418^R$ colonies from each strain were selected and screened by PCR in order to confirm correct cassette integration (Section 5.3.2b). The screens were designed to amplify across the two novel junctions created (Figure 5.16). For both strains (and for each colony screened) the expected sized PCR products were obtained: 675 b.p. for the LHS junction and 678 b.p. for the RHS junction, confirming successful integration of the *lox* flanked *kanMX4* module at the *LEU2* locus.

Southern blot analysis was performed on one transformant from each new strain (Section 5.3.2c). The oligonucleotide to be used as the probe, FL2B (Table 5.4), was end-labelled with ^{32}P by incubating with T4 polynucleotide kinase for 1 hour at 37°C. Percentage incorporation of the radioisotope into the probe was determined by scintillation count and found to be 20% (see Discussion). The Southern blot is presented in Figure 5.19. The size of the expected band is 1339 b.p., this is observed for both strains. A second fainter band is also seen for both strains at approximately 1.7 k.b.p. (shown to be a product of incomplete digestion, see

Discussion). No bands are observed, as expected, for the negative control lanes containing wild-type FY10 and FY22 genomic DNA.

The presence of single bands (excluding the incomplete digestion bands) at the correct marker position confirmed that no other copies of the *loxP-kanMX4-loxP 511 LEU2* sequence had been inadvertently integrated into the genomes of FY10 and FY22 and that the disruption cassette was correctly integrated at the *LEU2* locus in both strains.

Further, the integrity of each *lox* site in the disruption cassette of strain FY10 was verified by DNA sequencing (Section 5.3.2d). The PCR fragment amplified for sequencing purposes (the entire disruption cassette locus) was found, when analysed by agarose gel electrophoresis, to be in agreement with the expected size; 2549 b.p. Sequencing revealed both the *loxP* and *loxP 511* sites to be correct, and free from mutations. The sequencing also showed that the upstream *leu2* region contained two changes to the SGD sequence: a 5' CG 3' dinucleotide in the SGD sequence (coordinates 90740 - 90741) was found to be inverted.

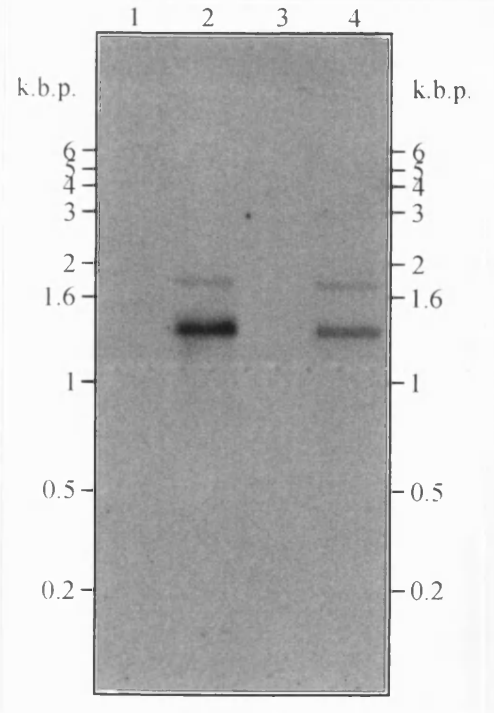


Figure 5.19 Southern blot analysis of strains FY10 (lane 2) and FY22 (lane 4) containing the *loxP-kanMX4-loxP 511 LEU2* gene disruption cassette. The band at 1.3 k.b.p. in both strains confirms the presence of the gene disruption cassette. FY10 only and FY22 only negative controls are shown in lanes 1 and 3 respectively.

5.5.3 Construction of targeting vector pRS315[*OR17-228*]

The open reading frame of the putative human odorant receptor, *OR17-228*, was amplified (Section 5.3.3b) from a previously made construct, the *HML* disruption cassette (this was to allow the recognition site for the rare cutter endonuclease I-*Sce*I to be incorporated at the 3'-end of the *OR17-228* ORF, a requirement for other work). Incorporation of the I-*Sce*I restriction site resulted in the fragment containing 20 n.t. of sequence (immediately after the I-*Sce*I site) homologous to the *LEU2* gene ORF (see Discussion). The size of the PCR product when analysed by agarose gel electrophoresis was in accordance with that expected; 1022 b.p.

The fragment was double-digested with restriction enzymes *Sph*I and *Xba*I (Section 5.3.3c) and ligated into vector pRS315-*lox* D1.3 bearing compatible cohesive ends, resulting in the *OR17-228* ORF being flanked by a *loxP* site and a *loxP 511* site. The ligation was transformed directly into *E. coli* cells. Five ampicillin resistant colonies were PCR screened and all were found to contain the insert, each producing a fragment corresponding with that expected; 1289 b.p. The construct from one of these colonies was partially sequenced and the integrity of both *lox* sites verified (the I-*Sce*I restriction site was also shown to be free from mutations; also the same three mutations at the 3'-end of the *OR17-228* ORF as observed previously [Table 3.14] were again noted).

The new targeting vector containing the *lox* flanked *OR17-228* ORF was designated pRS315[*OR17-228*]; the orientation of the *OR17-228* ORF with respect to the *lox* sites in the vector is shown in Figure 5.20.

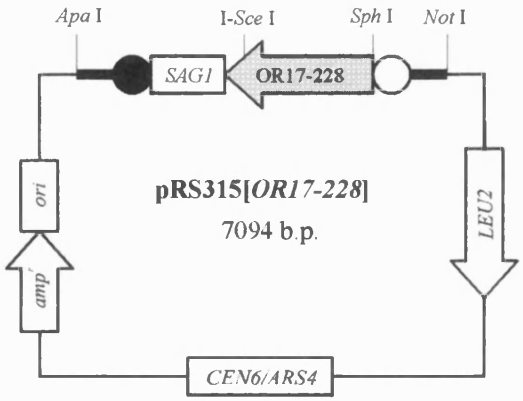


Figure 5.20 The Cre-recombinase targeting vector pRS315[*OR17-228*]. Closed and open circles represent *loxP* and *loxP 511* sites respectively. Other features are as described for vector pRS315-*lox* D 1.3 (Section 5.2.2b). Figure is not to scale.

5.5.4 Transformation of the *LEU2* disrupted FY10 strain with pBS39 and pRS315[*OR17-228*]

The Cre-recombinase expression vector pBS39 and the targeting vector pRS315[*OR17-228*] were co-transformed (Section 5.3.4) by electroporation into the FY10 strain containing the *loxP-kanMX4-loxP 511* cassette integrated at the *LEU2* locus (BDY1; Section 5.2.1). Only 2 Ura^+ Leu^+ transformants were obtained; however, when PCR screened, each colony was shown to contain both vectors. The new strain was designated BDY3.

The individual transformation of vector pBS39 into BDY1 yielded a far greater number of transformants (~200 colonies/plate), 2 colonies were screened and both were shown to contain the vector. The new strain was designated BDY4.

5.6 RESULTS: CRE-MEDIATED RECOMBINATION

The strain containing the recombination model system, BDY3, was grown in liquid medium containing galactose for 24 hours in order to induce Cre expression and promote *lox* site recombination (Section 5.4.1). After the induction period the cells were grown on SC-Leu medium from which 100 colonies were selected and transferred onto a gridded YPD plate containing G418, in order to identify G418-sensitive ($G418^S$) colonies. Of the 100 colonies picked, 35 failed to grow.

Recombination events in the 100 colonies (from the original SC-Leu plate) were examined by PCR screening (Section 5.4.2a). The various screening bands that were obtained are shown in Figure 5.21.

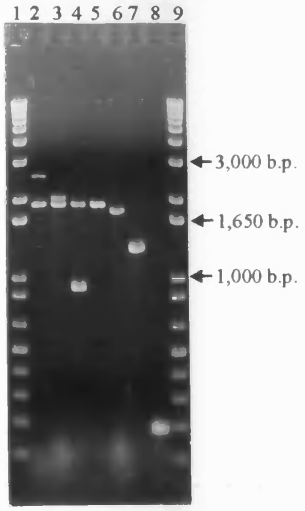


Figure 5.21 1% agarose gel showing the different bands obtained from PCR-screening of recombination induced BDY3 cells. Screening of the DNA element between the *lox* sites at the *leu2* locus produced the bands shown in lanes 2 - 5. Note that the second band observed in lanes 2 - 4 (1930 b.p.) corresponds to the *LEU2* locus amplified from the vector pRS315[*OR17-228*] (see Discussion). Lanes 1 and 9, the 1 k.b.p. DNA marker (Figure 2.1). Lane 2, the *kanMX4* module (2549 b.p.). Lane 3, the *OR17-228* ORF (2036 b.p.). Lane 4, deletion of the DNA element (939 b.p.). Lane 5, recombination intermediate (1930 b.p.). Screening of the DNA element between the *lox* sites of the targeting vector produced the bands shown in lanes 6 - 8. Lane 6, the *kanMX4* module (1778 b.p.). Lane 7, the *OR17-228* ORF (1266 b.p.). Lane 8, deletion of the DNA element (194 b.p.).

Of the 35 G418^S colonies sixteen indicated the presence of the *OR17-228* ORF (lane 3, Figure 5.21). Partial sequences of the *leu2* loci from 5 of these colonies showed that the exchange had occurred as intended (Figure 5.18), with the *OR17-228* ORF placed in the desired orientation, and each recombined *lox* site correctly formed. The five colonies were further analysed by Southern blotting in order to confirm correct integration of just one copy of the target ORF. Sauer (1992) had identified 10 cryptic *lox* sites in the yeast genome. However, hybridisation with a ^{32}P-end labelled oligonucleotide, OR17-228(F), showed that for all 5 colonies, no other copies of the *OR17-228* ORF had been inadvertently integrated, each colony producing the expected single band at 2281 b.p. (Figure 5.22).

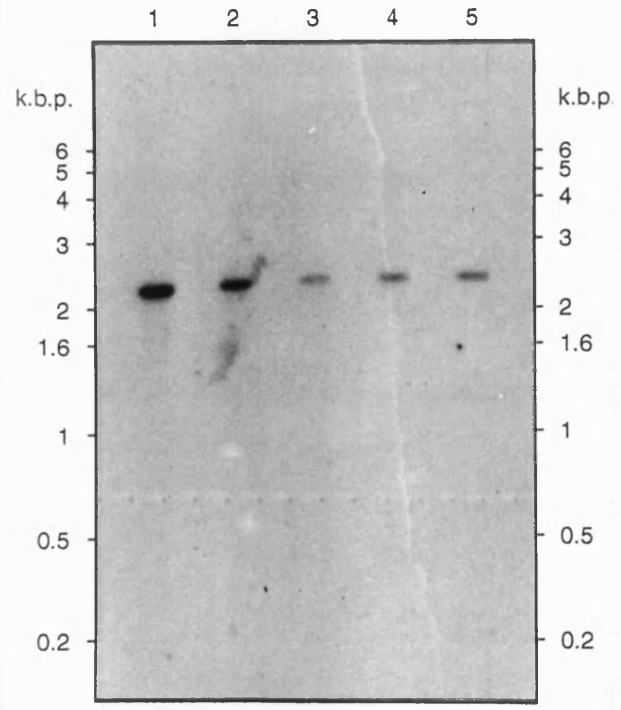


Figure 5.22 Southern blot analysis of 5 colonies showing the desired exchange event. A single band at 2.3 k.b.p. reveals that *OR17-228* ORF has been correctly targeted.

For 16 of the other 19 $G418^S$ colonies, the PCR screen indicated that the chromosomal DNA element lying between the two *loxP* sites, either the *kanMX4* module or the *OR17-228* ORF, had been excised (lane 4, Figure 5.21). Partial sequences of the *leu2* loci from three of these colonies (Section 5.4.3) showed that excision has occurred to leave just a single *lox* site, presumably through recombination of the *loxP* and *loxP 511* sites (1 sequence corresponded to *loxP* and 2 to *loxP 511*). The remaining 3 $G418^S$ colonies represented intermediates in the exchange process, producing the bands shown in lane 5, Figure 5.21. The nature of these intermediates was further investigated (Section 5.7.2), see Discussion.

PCR screening of the 65 $G418^R$ colonies revealed only six colonies in which the *kanMX4* module had been exchanged for the *OR17-228* ORF. The majority (fifty-one) of $G418^R$ colonies retained the chromosomal copy of the *kanMX4* module. The 8 remaining colonies were $G418^R$ as a result of an episomal copy of the *kanMX4* module; the *leu2* locus PCR screen for these colonies indicated a lack of a *loxP*-flanked element (see Discussion). Of the 51 $G418^R$ colonies with a chromosomal copy of the *kanMX4* module, about half (twenty-three) contained pRS315 plasmid in which the *OR17-228* ORF had been lost. The complete set of results of the PCR analysis of the 100 colonies is given in Table 5.8.

In the control for the above experiment, where BDY3 was grown in glucose-containing medium (in which Cre recombinase production should be suppressed), the colonies picked from the SC-Leu plate to the G418-containing YPD plate were all G418-resistant, as expected.

The above experiment was repeated several times, with similar results (see Discussion). However, the above experiment represents the most complete analysis.

Using strain BDY4 (Section 5.2.1), Cre induction was repeated in the absence of the targeting vector [Note that the method was as given in Section 5.4.2 for induction of strain BDY3, except that the selective growth medium was SC-Ura instead of SC-Ura-Leu]. 26/100 $G418^S$ colonies were produced (no $G418^S$ colonies were produced in the glucose control). PCR screening and partial sequences of the *leu2* locus of three colonies showed that the $G418^S$ colonies had arisen due to excision of the *loxP*-flanked *kanMX4* module (2 sequences corresponded to *loxP* and 1 to *loxP 511*).

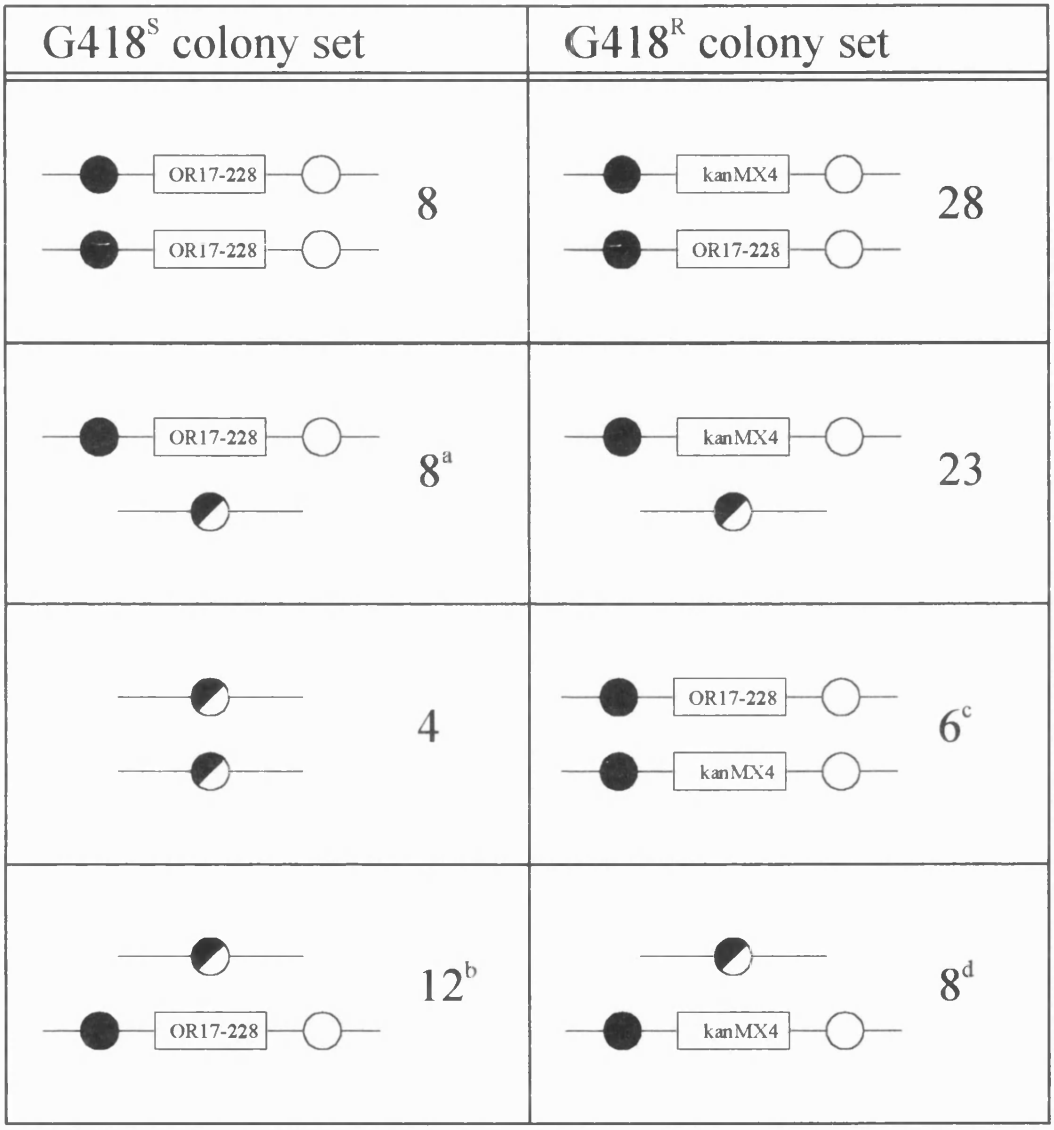


Table 5.8 Results of the PCR analysis of 100 colonies following Cre expression. In each case the top line represents the DNA element present between the *lox* sites at the *leu2* locus and the bottom line the DNA element found between the *lox* sites in the targeting vector. *loxP* site (filled circle), *loxP 511* site (empty circle). The PCR screen of the vector for some colonies produced an additional band or bands to the one shown (see Discussion), thus: (a) 3 colonies showed the *OR17-228* ORF; (b) 3 colonies showed intramolecular recombination; (c) 2 colonies showed the *OR17-228* ORF/1 colony the *OR17-228* ORF and intramolecular recombination; (d) 1 colony showed the *OR17-228* ORF and intramolecular recombination/1 colony showed intramolecular recombination. Three colonies, not represented above, were intermediates in the exchange process.

Intramolecular recombination between *loxP* and *loxP 511* sites was also investigated at a different chromosomal locus, *SAG1*, using strain CAY1 (Section 5.2.1). In CAY1 the *SAG1* locus has been disrupted with the *loxP-kanMX4-loxP 511* cassette. CAY1::pBS39 was induced as for strain BDY4 (above) and the number of $G418^S$ colonies determined. In two separate experiments 13/100 and 18/100 $G418^S$ colonies were produced (for the first experiment 1 $G418^S$ colony was observed for the glucose control, and none for the second experiment).

5.7 DISCUSSION

5.7.1 Generation of experimental strains for the analysis of Cre-mediated recombination

The *LEU2* gene disruption cassette and the targeting vector pRS315[*OR17-228*] were successfully made using routine recombinant methods. In both cases the *lox* sites that flank the *kanMX4* module and the *OR17-228* ORF were supplied by the vector pRS315-*lox* D1.3. The *loxP* and *loxP 511* sequences are directly repeated with respect to each other (discussed in Section 5.1.1), their orientation in the constructs made is shown in Figure 5.23.

loxP 5' ATAACTTCGTATA GCATACAT TATACGAAGTTAT 3'
loxP 511 5' ATAACTTCGTATA GTATACAT TATACGAAGTTAT 3'

Figure 5.23 Orientation of the *lox* sites in the *LEU2* gene disruption cassette and the pRS315[*OR17-228*] targeting vector. Note that the sequences are inverted with respect to Figure 1 of Hoess *et al.* (1986) [reproduced in part in Figure 5.2].

Successful integration of the disruption cassette at the *LEU2* locus of two FY-strains was confirmed by PCR screening of the novel joints created upon integration. It was further shown by means of Southern blot hybridisation, using a ^{32}P end-labelled oligonucleotide probe, that only one copy of the *lox* flanked *kanMX4* module had been integrated into the genomes of both strains, and that no illegitimate copies were present. The new strains were designated BDY1 and BDY2 (Section 5.2.1).

The targeting vector pRS315[*OR17-228*] and the Cre-recombinase expression vector pBS39 were successfully co-transformed into strain BDY1 (strain BDY2 was used by S. Aijaz for separate work). All the *lox* sites present in BDY1 were verified by DNA sequencing and shown to be correct. The new strain containing the completed model system (Section 5.1.2), was designated BDY3 (Section 5.2.1). Additionally, in order to study recombination events in the absence of the targeting vector, strain BDY4 (containing only pBS39) was created.

General considerations

Genomic DNA for the Southern blot analysis was digested with the restriction endonuclease *Ear*I, the position of the relevant cut sites is shown in Figure 5.16. The choice of an enzyme that cleaves in the middle (approximately) of the target sequence was not ideal, as any misintegration events at the 3'-end of the cassette would not be revealed. The choice of an enzyme that only cleaved externally to the integrated fragment would have overcome this problem (e.g. *Hpa*I, as used for subsequent analysis of exchange events).

The Southern blot hybridisation probe was generated by means of a kinase reaction in which $^{32}P_i$ was incorporated into the primer FL2B. It was found by means of a scintillation count that the percentage incorporation of the radionucleotide into the primer, after 1 hour incubation with approximately 100× excess of T4 polynucleotide kinase, was only 20%. Low sensitivity and indistinct bands were to prove a problem in later Southern blots; better end-labelling of the probe would undoubtedly have increased sensitivity and generated more reproducible results.

The Southern blot shown (Figure 5.19) revealed, as well as the expected bands at 1339 b.p., second fainter bands at 1.7 k.b.p. A preliminary Southern blot is shown in Figure 5.24 (sample preparation and conditions were as previously) in which the amount of enzyme required to achieve complete digestion was investigated. It is clearly seen that as the amount of enzyme increases (left to right) the primary band becomes more intense and the secondary bands vanish. The secondary bands observed are therefore unquestionably due to incomplete sample digestion. It is unfortunate that the samples for the actual Southern blot failed to digest to completion; a second blot (not shown) revealed no secondary bands but had run too far.

Between the *lox* sites of the targeting vector is the DNA sequence intended for exchange via Cre-mediated recombination. This sequence (primarily the *OR17-228* ORF) contains a sequence, 20 n.t. in length, homologous to the intended target, the *LEU2* locus. The presence of this sequence might be judged undesirable as it could potentially interfere with recombination events by providing an alternative pathway, namely DNA repair by homologous recombination. However, this is most unlikely as the sequence is very short and surrounded by sequence heterologous to the *LEU2* locus. Further, and perhaps more crucially, if recombination were to proceed as

expected (Section 5.1.2) then the sequence would be in the wrong orientation with respect to the *LEU2* sequence in order to recombine.

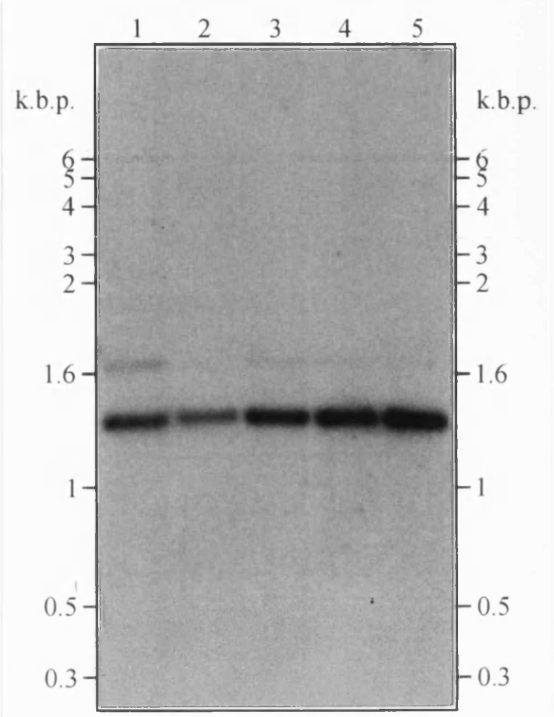


Figure 5.24 Preliminary Southern blot analysis to investigate the amount of restriction endonuclease (*Ear*I) required for complete sample digestion. Lanes 1 to 5 correspond to 10, 20, 30, 40 and 50 units of enzyme respectively (in a 100 µl reaction volume).

5.7.2 Cre-mediated recombination

General considerations

PCR screening of the targeting vector occasionally resulted in two (and in two cases three) different bands being amplified from one colony (see Table 5.8). The targeting vector, although a low copy number plasmid (containing the *ARS* and *CEN* sequence elements), is present between 1 and 5 copies per cell (R&D Systems pYX131 Catalogue). The double/triple bands observed were therefore as might be expected, each one being amplified from a separate vector, with each copy containing a different *lox*-flanked (or deleted) DNA element (the nature of the recombination events observed is discussed further in the next section). In the sixteen $G418^S$ colonies indicating the presence of the *OR17-228* ORF at their *leu2* loci, the existence of more

than one copy of the vector was again evident, as 8 of the colonies also still maintained an episomal copy of the sequence (Table 5.8). Given that the chromosomal element must have arisen by reciprocal exchange of sequences between a targeting vector and the *LEU2* locus (followed by loss of the *kanMX4* element from the vector, i.e. the colonies are $G418^S$), then it would be expected that the eight colonies showing a vector-borne copy of the *OR17-228* ORF should also show a targeting vector with a deleted DNA element, unusually however, no such vector is observed.

To prevent the vector-borne copy of the *OR17-228* ORF sequence interfering with the Southern blot, the targeting vector was outgrown in each of the 5 colonies analysed. This was successfully done by repeated culturing (3 times) in non-selective medium (Section 5.4.4), after which it was found that approximately 95% of cells had lost the vector and were Leu^-. Genomic DNA for Southern blot analysis was then prepared from these colonies.

As previously mentioned, the sensitivity of the Southern blot was to prove troublesome. In order to obtain bands of sufficient clarity it was necessary to combine genomic DNA preparations, and to run the samples in as small a volume as possible. The choice of autoradiography film was to prove important, with the best results obtained using X-Omat AR (Sigma). As already stated, sensitivity might have been improved by better end-labelling of the oligonucleotide probe. However, percentage incorporation of ^{32}P into FL2B was not determined here (in order to avoid any unnecessary manipulation of the radioactive material); the kinase reaction was added directly, without removal of unincorporated [γ-^{32}P]dATP, to the hybridisation buffer. This did not appear to result in any increase in the observed background.

PCR screening of the *lox*-flanked DNA element at the *leu2* locus had revealed an unexpected secondary band at 1930 b.p. (Figure 5.21; lanes 2 - 4). This band was an unfortunate consequence of having chosen a *leu2*Δ-based target in conjunction with a *LEU2* targeting vector; the primers designed to amplify the target locus also amplified the *LEU2* sequence from the vector! The band, although an undesirable competing reaction in the PCR screens, was fortunately easily resolved from those of interest.

The three intermediates observed in the exchange process also resulted from the *leu2*Δ-target/*LEU2*-vector choice. These $G418^S$ colonies produced no band for the vector screens, and a single band (corresponding in size to the *LEU2* band from the

vector) for the *leu2* locus screen (lane 5, Figure 5.21). These colonies, which remained Leu^+ after repeated growth in non-selective medium, appeared to be the result of either gene conversion of the *leu2* locus (with concomitant loss of the targeting vector), or permanent integration of the targeting vector at the *leu2* locus.

The exact nature of the intermediates was investigated using the primer pair GCvCO and Leu2TB4 (Section 5.4.4c). The PCR analysis showed that the intermediates were the result of several recombination events. With reference to Figure 5.25: intermediate A has undergone homologous recombination between episomal *LEU2* and the chromosomal *leu2* DNA, leading to excision of the *kanMX4* module and the formation of structure P (confirmed for each colony by the production of the expected size band; 2330 b.p.). Further PCR screening (using primer pairs Leu2PF3 (Table 5.4)/T7 (Table 5.2) and OR17-228(F)/OR17-228(B) [Table 3.5]) revealed that in two of the colonies an additional recombination event between the *loxP* and *loxP 511* sites had resulted in removal of the intervening element, leading to the formation of structure Q. Structure Q could also have arisen from the intermediate formed by intermolecular recombination between *loxP 511* sites (structure B, Figure 5.7). In this eventuality the excised intervening element would have been the *kanMX4* module.

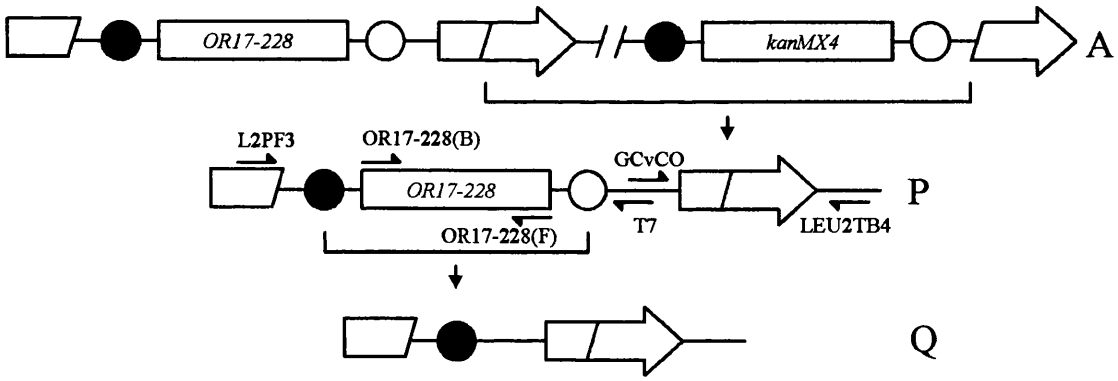


Figure 5.25 Recombination intermediates. Structure **A** is derived from recombination between *loxP* sites (see Figure 5.7). Broken arrows: chromosomal and episomal *LEU2* DNA. Structures **P** and **Q** are as described in the text. The positions of the various primers used to determine the nature of the intermediates are shown (Figure is not to scale).

Efficiency of Cre-mediated exchange

It was found that just over 20% of cells exhibited the desired targeting event (Table 5.8). This compares quite favourably with the 50% that would be expected if a 1:1 equilibrium had been established between the exchanging species (Figure 5.7). Waterhouse *et al.* (1993), in a bacterial Cre-*lox* system utilising the same *lox* site combination as this study, noted a 3:1 bias in favour of the desired targeting event. This was attributed to differences in the efficiency of inter- and intramolecular recombination at the *loxP* and *loxP 511* sites (it is unclear, however, how such differences would result in any bias towards the targeting event; although vector-based the system is analogous to that shown in Figure 5.7, where any bias would still result in a 1:1 equilibrium). A further 10% (approximately) of cells appear to have undergone exchange, but then subsequently lost the transgene through intramolecular *lox* site recombination. Amongst those cells that had retained the antibiotic marker (roughly 50% of the population), about half possessed a targeting vector from which the donor gene had been excised.

Ross-MacDonald *et al.* (1997) reported greater than 90% intramolecular recombination between the highly divergent *loxP* and *loxR* sites (Section 5.1.2). This suggested that the more similar *loxP* and *loxP 511* sites might recombine at an even higher frequency, and, as previously suggested, render the double recombination gene targeting strategy useless. Although the above results showed intramolecular recombination between the *loxP* and *loxP 511* sites to be a problem, the system works a lot better than expected. This is likely to be due to the unexpectedly low efficiency of intramolecular recombination observed between the *loxP* and *loxP 511* sites (~25% at the *LEU2* locus and between 15 and 20% at the *SAG1* locus) and the competing intermolecular and intramolecular recombination between identical *lox* sites.

In this experiment 47% of the $G418^S$ colonies were found to be exchangeants, compared to 9% amongst the $G418^R$ colonies. Another experiment (in which 50 colonies were PCR screened after Cre induction) revealed 31% of the $G418^S$ colonies to be exchangeants and none of the $G418^R$ colonies. This increased number of exchangeants amongst $G418^S$ colonies was also observed in an experiment where Cre-expression was halted after 1 h and 5 h induction (here however, only 15 colonies were screened at each time point). After 1 h induction 25% of $G418^S$ colonies were exchangeants and 10% of the $G418^R$ colonies; after 5 hours induction 57% of $G418^S$ colonies were exchangeants and none of the $G418^R$ colonies. It was reasoned that it

might be advantageous to increase the induction period of Cre expression beyond 24 h. It was found that increasing the induction period did increase the number of $G418^S$ colonies (1 h: 0-4/100; 5 h: 1-7/100; 12 h: 7-8/100; 24 h: 28-37/100; 36 h: 34-36/100; 48 h: 32-45/100; ranges quoted are the results of 3-6 experiments), although no $G418^S$ colonies from experiments involving induction periods longer than 24 h were analysed in detail.

The results from several different experiments have been reported, those being: Cre-mediated exchange following 24 h galactose induction (strain BDY3); exchange following 1 h - 48 h galactose induction (above; strain BDY3); and 24 h galactose induction for *lox* site recombination in strains lacking a targeting vector (BDY4 and CAY1). For each of these experiments, a control was performed, where the same number of colonies (as for their galactose-induced counterpart) were plated on G418 medium following induction in glucose-containing medium. As expected, no recombination was observed for the control experiments (no $G418^S$ colonies were produced). For one experiment - 24 h induction of CAY1 - 1 $G418^S$ colony was observed for the glucose control, in a second repeat experiment no $G418^S$ colonies were observed. Importantly, therefore (for the stability of any integrated library), recombination between the two *lox* sites appears to be dependent upon the presence of the Cre recombinase, with spontaneous recombination (or recombination induced by transient expression of Cre) occurring at a maximum frequency of $\sim 10^{-3}$. It is likely that the actual frequency of recombination - in glucose-induced cells - is much lower than this. If the average recombination frequency is calculated from several similar 24 h induction experiments (data not shown), then the spontaneous recombination frequency is probably $<10^{-4}$.

Overall efficiency of the double cross-over strategy

The maximum transformation efficiency achievable by standard integration methods is about 10^3 transformants per μg (Hill, 1989; Rothstein, 1991). In comparison, the maximum efficiency of episomal vector transformation - upon which the double cross-over strategy is based - is about 10^6 transformant per μg (Manivasakam and Schiestl, 1993). Potentially, the integration efficiency achievable with the Cre-*lox* system in this context, assuming a 1:1 equilibrium between the exchanging species, is $\sim 10^5$ transformants per μg. Although the efficiency of the system reported here was reduced due to the competing intramolecular recombination (as discussed above), it

remains about two orders of magnitude better than existing integration methods, such as PCR-mediated direct gene replacement (Wach, 1996).

It is acknowledged, however, that certain additional steps are required in the double cross-over system compared to, e.g., PCR-mediated direct gene replacement. Firstly, it is (of course) required that the recipient strain can express the Cre protein (in a controllable manner), requiring an initial transformation with a suitable vector. Secondly, following transformation with the targeting vector (containing the exogenous DNA to be integrated), the transformants must be induced for the exchange event (no reduction in cell viability was observed following induction, with glucose- and galactose-induced cells plating with the same efficiency onto SC-Leu medium). Other steps between the two processes remain comparable: Both require an initial transformation with a PCR-generated construct, containing a selectable or counterselectable marker, to target the genomic locus of interest; and secondly, constructs (if considering a library) containing the transgenes must be prepared. In PCR-mediated direct gene replacement, this entails adding regions of flanking homology - by PCR - to the library (the library might represent, for example, shuffled genes in a molecular evolution strategy). For Cre-mediated exchange, the parallel step would be to ligate the library of shuffled genes into a suitably digested targeting vector.

Although the model system was tested with construction of stable DNA repertoires in mind, it was also found to be very reliable in routine integrations (Adams, 1999). Its reliability stems primarily from the initial selection for antibiotic-resistant yeast, which is straightforward. PCR screening of 10-20 $G418^S$ colonies following Cre expression and replica plating onto G418-containing medium always results in a colony which exhibits the desired integration event. The main disadvantage is the need to initially make, and then integrate, the antibiotic resistance construct.

Possible ways of improving the Cre-lox double recombination strategy in yeast

If the double cross-over system were to be used as described here, it would be advisable to attach a selectable marker to the DNA elements being integrated, to facilitate identification of the desired integrants. It would also be advisable to remove the targeting vector by several rounds of growth in non-selective medium following Cre expression. However, if the efficiency of exchange could be improved then these

additional steps could be avoided. One obvious way in which to improve the Cre-*lox* double recombination system would be to find a pair of *lox* sites which recombine very poorly in yeast (or better still not at all). Combinations of known *lox* sites, both wild-type and mutant could be tried in a variety of different orientations. Albert *et al.* (1995) investigated the possibility of using pairs of *lox* sites with mutated inverted repeats to drive the Cre-catalysed reaction in favour of the desired integration event (Figure 5.26).

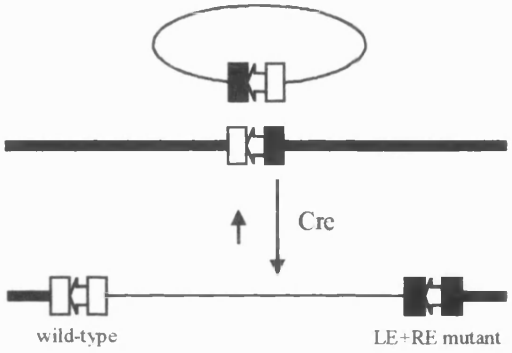


Figure 5.26 Site-specific recombination between a *lox* site on a plasmid and a *lox* site on a chromosome inserts the circular molecule into the chromosome. The *lox* site is depicted as a tripartite structure with an asymmetric 8 b.p. spacer (arrowhead) and two 13 b.p. elements (rectangles). Filled rectangles represent mutant elements. Recombination between a LE (left element) mutant site and a RE (right element) mutant site produces a wild-type and LE+RE mutant site. If the LE+RE mutant site is poorly recognised by Cre, then the inserted molecule is less likely to be excised (Albert *et al.*, 1995).

This is an interesting strategy for improving the stability of integrated elements that are formed via a single intermolecular cross-over event (such elements are unstable as they are usually flanked by identical *lox* sites, intramolecular excision of which is kinetically favoured). Araki and co-workers (1997) showed that this strategy could be successfully used to increase the targeting efficiency in embryonic stem cells, integration frequency via mutant *lox* sites reached a maximum of 16%, compared to <0.5% between wild-type *loxP* sites. Unfortunately however, due to the more complex nature of the intermediates involved in the double cross-over system (A and B, Figure 5.7), the implementation of a similar approach to drive product formation might prove to be very difficult indeed. For such a strategy to work it would be necessary not only to have a pair of *lox* sites that do not recombine with one another, e.g. the hypothetical sites *loxM* and *loxN*, but also left and right element

mutants of each site, i.e. *loxM*(LE)/*loxM*(RE) and *loxN*(LE)/*loxN*(RE), that only recombine with their respective partners. To find such a combination of *lox* sites might prove impossible.

Finally, another possible approach would be to mutate the *cre* gene in order to increase specificity[Ψ]. It was suggests by Kwon *et al.* (1997), studying the catalytic core of the related lambda integrase, that alterations in the flexible delivery arm of catalytic residue Tyr-324 (which plays a critical role in DNA binding and site specificity amongst the Int family of recombinases) could be expected to significantly influence such parameters as the rate, reversibility, and specificity of DNA cleavage and ligation reactions. As part of a study investigating the amino acids involved in Cre-DNA recognition, Hartung and Kisters-Woike (1998) mutated an amino acid adjacent to the catalytic tyrosine residue: Asn-323→Arg-323. In this example however, the mutation resulted in a broadening of the recombination specificity.

[Ψ][Note added during corrections: The feasibility of altering the Cre recombinase has recently been demonstrated by Buchholz and Stewart (2001), who used a directed evolution strategy to successfully alter Cre's site specificity.]

5.7.3 Conclusions

The Cre-*lox* double recombination strategy has been successfully applied in other organisms (Section 5.1.2). The results presented here indicate that the strategy, with some refinement, could be an equally effective, and very useful tool, for genome engineering in yeast. However, some care may be needed before applying the system to a real situation. For instance, the *lox* sites might interfere with the expression of an ORF which they flank. In eukaryotes, translation initiation is believed to follow a scanning mechanism, whereby the 40S ribosomal subunit plus various co-factors bind to the 5' cap of the mRNA, then translate down the untranslated leader scanning for the first AUG codon (Kozak, 1999). In yeast the mRNA leader sequence has an average length of about 50 n.t. (Romanos *et al.*, 1992). The most significant factor causing inhibition of translation is secondary structure formation within the leader sequence. A *lox* site, possessing two inverted repeats, has the potential to form just such inhibitory secondary structure (Figure 5.27). It would therefore be necessary to position the *lox* sites carefully in order to avoid this possibility. However, Fukushige and Sauer (1992), investigating Cre-*lox* genomic targeting in mammalian CHO cells,

fused an ATG-*loxP* sites to their positive selection gene *neo* (removing the first 5 codons of the *neo* gene which had previously been shown to be nonessential). A functional *lox-neo* fusion protein was produced which rendered the CHO cells resistant to G418 (which is an analogue of neomycin). It is unclear why such a strategy does not succumb to the above mentioned pitfalls of secondary structure inhibition, it is nevertheless interesting and encouraging that an in-frame fusion resulted in a functional protein. A similar strategy was also successfully used by O'Gorman *et al.* (1991) who created a fusion protein between the yeast FLP recombinase recognition site, FRT, and the bacterial β-galactosidase coding sequence, producing an ATG-FRTβGAL fusion that showed good activity in monkey kidney cells and mouse embryonal carcinoma cells.

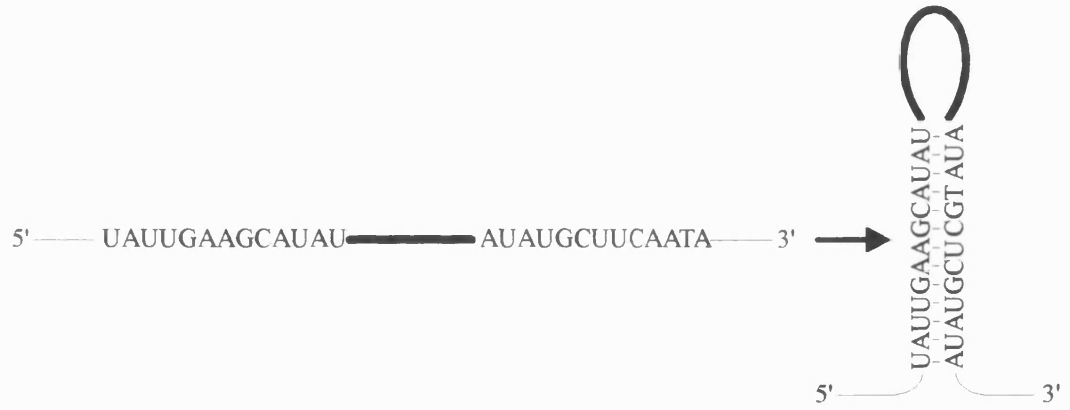


Figure 5.27 The *lox* site as part of a mRNA leader sequence (thin lines: leader sequence; thick line: 8 b.p. *lox* spacer region). The 13 b.p. inverted repeats have the potential to form a secondary hairpin structure which may inhibit efficient expression of a heterologous ORF.

Finally, another potential problem is the fact that the *lox* sites are introduced and remain in the genome. The integration of further constructs with the Cre-*lox* system would need to be done using different *lox* sites from those already present in order to prevent misintegration at the original site and the possibility of chromosomal translocations.

The above mentioned problems notwithstanding, the possibility of developing the Cre-*lox* double recombination strategy further in the yeast is an attractive possibility. Immediate future work would be to investigate recombination between various heterologous *lox* sites. As demonstrated by Ross-Macdonald *et al.* using the

loxP, loxR system, and here using the *loxP, loxP 511* combination, Cre-mediated recombination might not be as straightforward and as clear-cut in yeast as in other systems, in that dissimilar sites recombine. An important question is thus posed: why is heterogeneity in the spacer region not sufficient to prevent unsolicited *lox* recombination in *Saccharomyces cerevisiae*?

CHAPTER 6

GENERAL DISCUSSION

6.1 Directed molecular evolution by gene conversion

The projects aims, from an experimental perspective, were: **a**. the construction of an experimental strain containing the synthetic donor and acceptor loci (containing the heterologous genes to be shuffled) and a vector capable of expressing the I-*Sce*I endonuclease in a controllable manner (as described in Section 1.3); and **b**., to study recombination events following induction of gene conversion (by galactose induced expression of I-*Sce*I). For the reasons described previously (Section 1.3), gene conversion events were also to be monitored in strains defective for mismatch repair. If gene conversion was shown to occur as intended (in either a wild-type or mismatch repair defective strain) then gene conversion events in strains carrying more than one donor gene could be investigated.

Two experimental strains were successfully constructed; one wild-type and one containing a deletion of the mismatch repair gene, *PMS1*. Additionally, two control strains, one containing no donor locus and the second containing a plasmid borne I-*Sce*I recognition site, were also successfully made. In each case a combination of PCR-screening and DNA sequencing was used to show that the desired elements were present and correct. However, the time taken to create the strains was too long, and, in the end, only allowed the briefest study of recombination events in either the experimental or control strains. With only enough time to perform each experiment once or twice, it was then very disappointing not to observe any evidence of the desired gene conversion reaction in the experimental strains, or evidence that I-*Sce*I was being expressed and cleaving its recognition site in either of the control strains.

The necessity to convert the experimental strain to a Gal^+ phenotype was perhaps the biggest stumbling block in the route taken to the construction of the experimental strain. It would undoubtedly have been a much better choice to have started with a Gal^+ strain, thus avoiding the time-consuming steps of mating,

sporulation and subsequent spore analysis. The parental strain, JRY50, was also *his3*, requiring another step to convert the strain to a His^+ phenotype.

JRY50 was chosen originally as it contained the temperature sensitive *sir4-9* mutation, allowing derepression of the donor locus (*HML*), a requirement for the direct gene replacement strategy, the first method attempted to integrate the donor gene (*OR17-228*) at *HML* (Section 3.3.3). Although this strategy failed (Section 3.4.6), and the ability to derepress the *HML* locus apparently was not critical for the second strategy adopted (although this aspect was not investigated), the strain was continued in its use, as the acceptor gene (*OR17-40*), had already been integrated (after considerable effort) at the *LEU2* locus. With hindsight, it might have been prudent at this stage to have changed to a more suitable strain, as, as well as being Gal^- His^-, the presence of the *sir4-9* mutation in the completed experimental strain is an unwanted, and potentially troublesome, complication.

Evidence from the preliminary experiments performed strongly suggests a problem relating to expression of the I-*Sce*I endonuclease. DNA sequencing was used to show that the sub-cloned I-*Sce*I ORF (from vector pPEX7 into pYES2) was correct, and the presence of the pYES2[I-*Sce*I] expression vector in all the strains made was confirmed by PCR screening. As suggested previously (Section 4.4.1), a Northern blot to look for evidence of expression of the endonuclease might be one way to address this problem. Firstly however, the promoter and terminator regions of the vector need to be sequenced (areas not covered in the sequencing already performed) to look for any deleterious mutations; alternatively the I-*Sce*I ORF might be cloned into another expression vector, and the experiments repeated. It would also be valuable to be able to perform the control experiments using the original I-*Sce*I expression vector, pPEX7. Attempts to create a 'vector control' strain - using the yeast strain FY833 - were made without success (discussed in Section 4.4.1). Unfortunately, again, due to a shortage of time, no attempts were made to construct such a 'vector control' strain in an alternative strain. This is also an important avenue of investigation for any future work.

6.2 Towards a Cre-based site-specific recombination system

The aim of the Cre-*lox* work was to investigate the fidelity and efficacy of a double-recombination strategy as a means to introduce DNA constructs site-specifically into the yeast genome. This investigation was completed as intended.

It was shown that the transgene (*OR17-228*) was targeted to the intended locus (*leu2*) in a little over 20% of the cells induced to undergo the exchange. PCR screening and Southern blot analysis of several of the targeted loci showed that the transgene had integrated correctly, and that there were no ectopic copies present.

Ways in which the double-recombination strategy might be improved have already been discussed (Section 5.7.2). Immediate future work (as previously suggested) would be to investigate recombination between different combinations of heterologous *lox* sites, with the aim of finding a pair that poorly recombine with one another (in order that the efficiency of the double-recombination strategy might be improved). However, as it stands, the screening strategy employed to look for exchangeants is quite laborious and time consuming, involving picking of individual colonies from YPD to YPD G418 plates, followed by extensive PCR analysis. Using this method the number of colonies that can be effectively screened in each experiment is limited. A more effective and rapid method for screening for exchangeants is needed if better *lox*-site combinations are to be identified. One way this might be achieved is to use the white → red colour change when an *ade1 ade4* strain is converted to an *ade1 ADE4* phenotype (Ray *et al.*, 1988). Such a system could be implemented by substituting the *OR17-228* ORF in the targeting vector with the open reading frame of the *ADE4* gene. The *lox*-flanked *kanMX4* gene disruption cassette could be modified to include the *ADE4* promoter exterior to the appropriate (depending on *ADE4* orientation in targeting vector) *lox* site. The disruption cassette containing the *lox*-sites to be investigated would then be transformed into an appropriate *ade1 ade4* strain, in addition to the targeting vector and Cre-expression vector. Following induction, colonies that have undergone the desired exchange event could be easily identified by means of a colour change from white to red due to restoration of a functional *ADE4* gene (Figure 6.1).

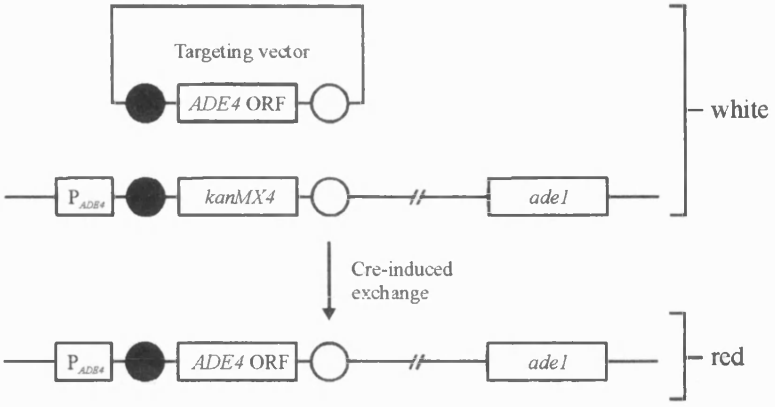


Figure 6.1 A possible mechanism by which exchangeants following Cre-mediated exchange might be identified, allowing better combinations of *lox* sites to be more easily identified without PCR screening. P_{ADE4} is the *ADE4* promoter, *lox* sites being investigated are shown as white and black circles. Correct targeted integration of the *ADE4* ORF into the genome results in restoration of the *ADE4* gene and an easily identifiable colour change from white to red due to the change in phenotype.

This strategy is similar to that devised by Call *et al.* (2000) for targeted integration in the mouse, whereby a vector containing a *lox* site, and a promoter for a gene conferring neomycin resistance, is targeted to a previously integrated *lox* site containing the promoterless neomycin resistance gene. Successful integration in this case leads to cells resistant to the antibiotic neomycin.

Using the white → red assay there would be no need to pick colonies from YPD to YPD G418 plates to identify $G418^S$ colonies (those that are most likely - in the present study - to have undergone exchange), or to PCR screen in order to identify exchangeants. In this way a large number of colonies could be screened for exchange events simultaneously, allowing the best combination of *lox* sites (and the best induction protocol) for the double-recombination strategy to be more easily identified.

REFERENCES

Abraham, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S. and Hicks, J. B. (1984). Regulation of mating-type information in yeast: negative control requiring sequences both 5' and 3' to the regulated region. *J. Mol. Biol.* **176**. 307 - 331.

Abremski, K., Hoess, R. and Sternberg, N. (1983). Studies of the properties of P1 site-specific recombination: Evidence for topologically unlinked products following recombination. *Cell.* **32**. 1301 - 1311.

Abremski, K. and Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259**. 1509 - 1514.

Adams, A., Gottscling, D. and Kaiser, C. (1996). *Methods in Yeast Genetics - A laboratory course manual.* Cold Spring Harbour Laboratory Press.

Adams, C. (1999). Strategies for producing therapeautic antibodies in yeast. Ph. D. Thesis. University of Bath.

Alani, E., Cao, L. and Kleckner, N. (1987). A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. *Genetics.* **116**. 541 - 545.

Albert, H., Dale, E. C., Lee, E. and Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *Plant Journal.* 7. 649 - 659.

Alt, F. W., Oltz, E. M., Young, F., Gorman, J., Taccioli, G. and Chen, J. (1992). VDJ recombination. *Immunol. Today.* **13**. 306 - 314.

Antoni, A. D. and Gallwitz, D. (2000). A novel multi-purpose cassette for repeated integrative epitope tagging of genes in *Saccharomyces cerevisiae*. *Gene*. **246**. 179 - 185.

Araki, K., Araki, M. and Yamamura, K. (1997). Targeted integration of DNA using mutant *lox* sites in embryonic stem cells. *Nucleic Acids Res.* **25**. 868 - 872.

Arnold, F. (1997). Directed evolution of enzyme catalysts. *Trends Biotech.* **15**. 523 - 530.

Arnold, F. (1998a). Design by directed evolution. *Acc. Chem. Res.* **31**. 125 - 131.

Arnold, F. (1998b). When blind is better: protein design by evolution. *Nature Biotechnology*. **16**. 617 - 618.

Astel, C. R., Ahlstrom-Jonasson, L. and Smith, M. (1981). The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*. *Cell*. **27**. 15 - 23.

Austin, S., Ziese, M. and Sternberg, N. (1981). A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell*. **25**. 729 - 736.

Barnes, G. and Rine., J. (1985). Regulated expression of endonuclease *Eco*RI in *Saccharomyces cerevisiae*: Nuclear entry and biological consequences. *Proc. Natl. Acad. Sci. USA*. **82**. 1354 - 1358.

Bailis, A. M. and Rothstein, R. (1990). A defect in mismatch repair in *Saccharomyces cerevisiae* stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics*. **126**. 535 - 547.

Ben-Arie, B., Lancet, D., Taylor, C., Khen, M., Walker, N., Ledbetter, D. H., Carrozzo, R., Patel, K., Sheer, D., Lehrach, H. and North, M. A. (1994). Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Hum. Mol. Genet.* **3**. 229 - 235.

Berry, D. R. (1982). Biology of yeasts. London: Edward Arnold.

Bethke, B. and Sauer, B. (1997). Segmental genomic replacement by Cre-mediated recombination: genotoxic stress activation of the p53 promoter in single-copy transformants. *Nuc. Acids. Res.* **25**. 2828 - 2834.

Bi, X. and Broach, J. R. (1997). DNA in transcriptionally silent chromatin assumes a distinct topology that is sensitive to cell cycle progression. *Mol. Cell. Biol.* **17**. 7077 - 7087.

Boeke, J. D., Lacroute, F. and Fink, J. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-floro-orotic acid resistance. *Mol. Gen. Genet.* **197**. 345 – 346.

Borst, P. and Greaves, D. R. (1987). Programmed gene rearrangements, altering gene expression. *Science*. **235**. 658 - 667.

Borts, R. H., Leung, W. Y., Kramer, W., Kramer, B., Williamson, S., Fogel, S. and Haber, J. E. (1990). Mismatch repair-induced meiotic recombination requires the *PMS1* gene product. *Genetics*. **124**. 573 - 584.

Boscheron, C., Maillet, L., Marcand, S., Tsai-Pflugfelder, M., Gasser, S. M. and Gilson, E. (1996). Cooperation at a distance between silencers and proto-silencers at the yeast *HML* locus. *EMBO Journal.* **15**. 2184 - 2195.

Brand, A. H., Breeden, L, Abraham, J., Sterglanz, R. and Nasmyth, K. (1985). Characterisation of a silencer in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell.* **41**. 41 - 48.

Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. and Broach, J. R. (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* 7. 592 - 604.

Brenneman, M., Gimble, F. S. and Wilson, J. H. (1996). Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site specific endonucleases. *Proc. Natl. Acad. Sci. USA*. **93**. 3608 - 3612.

Brown, W. S. (1976). A cross-over shunt model for alternate potentiation of yeast mating-type alleles. *J. Genet*. **62**. 81 - 91.

Buchholz, F. and Stewart, A. F. (2001). Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nature Biotechnology*. **19**. 1047 - 1052.

Buchman, A. R., Kimmerly, W. J., Rine, J. and Kornberg, R. D. (1988). Two DNA-binding factors recognise specific sequences at silencers, upstream activating sequences, autonomously replicating sequences and telomeres in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**. 210 - 225.

Buck, L. B. (2000). The molecular architecture of odour and pheromone sensing in mammals. *Cell*. **100**. 611 - 618.

Call, M. L., Moore, C. S., Stetten, G. and Gearhart, J. D. (2000). A Cre-*lox* recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. *Hum. Mol. Genet*. **9**. 1745 - 1751.

Campbell, I. and Duffus, J. H. [eds.] (1988). Yeast: A practical approach. Oxford IRL Press.

Carpenter, A. T. C. (1984). Meiotic roles of crossing-over and gene conversion. *Cold Spring Harbour Symp. Quant. Biol.* **49**. 23 - 29.

Chang, C. J., Chen, T. T., Cox., B. W., Dawes, G. N., Stemmer, W. P. C., Punnonen, J. and Patten, P. A. (1999). Evolution of a cytokine using DNA family shuffling. *Nature Biotechnology*. **17**. 793 - 797.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J. Lockhart, D. J., Moris, M. S. and Fodor, S. P. A. (1996). Accessing genetic information with high-density DNA arrays. *Science*. **274**. 610 - 613.

Chen, K. and Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA*. **90**. 5618 - 5622.

Chlebowicz, E. and Jachymczyk, W. J. (1979). Repair of MMS-induced double-strand breaks in haploid cells of *Saccharomyces cerevisiae*, which requires the presence of a duplicate genome. *Mol. Gen. Genet.* **167**. 279 - 286.

Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA*. **81**. 1991 – 1995.

Clackson, T. and Wells, J. A. (1994). *In vitro* selection from protein and peptide libraries. *Trends Biotech.* **12**. 173 - 184.

Clayton, R. A., White, O., Ketchum, K. A. and Venter, J. C. (1997). The first genome from the third domain of life. *Nature*. **387**. 459 - 462.

Clealand, J. and Craik, C. S. (1996). Protein engineering: Principles and practice. Wiley-Liss.

Coghlan, A. (1998). A sexual revolution. *New Scientist*. 21 November. 4.

Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A. and Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell*. **44**. 521 - 533.

Colleaux, L., d'Auriol, L., Galibert, F. and Dujon, B. (1988). Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc. Natl. Acad. Sci. USA*. **85**. 6022 - 6026.

Craig, N. L. (1988). The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **22**. 77 - 105.

Crameri, A., Cwirla, S. and Stemmer, W. P. C. (1996). Construction and evolution of antibody-phage libraries by DNA shuffling. *Nature Medicine*. **2**. 100 - 102.

Crameri, A., Dawes, G., Rodriguez, E., Silver, S. and Stemmer, W. P. C. (1997). Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nature Biotechnology*. **15**. 436 - 438.

Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*. **39**. 288 - 291.

Crameri, A., Whitehorn, E. A., Tate, E. and Stemmer, W. P. C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology*. **14**. 315 - 366.

Crickett, R. (1997). Microbiology with a sense of smell! B. Sc. Project Report. University of Bath.

Datta, A., Adouda, A., New, L., Crouse, G. F. and Jinks-Robertson, S. (1996). Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**. 1085 - 1093.

Dawes, I. W. and Hardie, I. D. (1974). Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether. *Mol. Gen. Genet.* **131**. 281 – 289.

Delneri, D., Tomlin, G. C., Wixon, J. L., Hutter, A., Sefton, M., Louis, E. J. and Oliver, S. G. (2000). Exploring redundancy in the yeast genome: an improved strategy for use of the cre-*lox*P system. *Gene*. **252**. 127 - 135.

Dietzel, C. and Kurjan, J. (1987). Pheromonal regulation and sequence of the *Saccharomyces cerevisiae SST2* gene: a model for desensitization to pheromone. *Mol. Cell. Biol.* **7**. 4169 - 4177.

Dryer, L. and Berghard, A. (1999). Odorant receptors: a plethora of G-protein-coupled receptors. *Trends Biol. Sci.* **20**. 413 - 417.

Dujon, B. (1996). The yeast genome project: what did we learn? *Trends In Genetics.* **12**. 263 - 270.

Fairhead, C. and Dujon, B. (1993). Consequences of unique double-stranded breaks in yeast chromosomes: death or homozygosis. *Mol. Gen. Genet.* **240**. 170 - 180.

Feldman, J. B., Hicks, J. B. and Broach, J. R. (1984). Identification of sites required for repression of a silent mating-type locus in yeast. *J. Mol. Biol.* **178**. 815 - 834.

Fishman-Lobell, J. and Haber, J. E. (1992). Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene *RAD1. Science.* **258**. 480 - 484.

Fishman-Lobell, J., Rudin, N. and Haber, J. E. (1992). Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.* **12**. 1292 - 1303.

Fukushige, S. and Sauer, B. (1992). Genomic targeting with a positive-selection *lox* integration vector allows highly reproducible gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA.* **89**. 7905 - 7909.

Galli, A. and Schiestl, R. H. (1998). Effects of DNA double-strand and single-strand breaks on intrachromosomal recombination events in cell-cycle-arrested yeast cells. *Genetics.* **149**. 1235 - 1250.

Gietz, R. D. and Woods, R. A. (1994). High efficiency transformation with lithium acetate in *Molecular Genetics of Yeast, A Practical Approach* (Johnston, J. R., ed.). 121 – 134. IRL Press, Oxford, UK.

Gilbert, W. (1978). Why genes in pieces? *Nature*. **271**. 501.

Giver, L., Gershenson, A., Freskgard, P and Arnold, F. (1998). Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. USA.* **95**. 12809 - 12813.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S. G. (1996). Life with 6000 genes. *Science*. **274**. 546 - 567.

Goffeau, A., *et al.* (1997). The yeast genome directory. *Nature*. **387** (suppl. 1). 1 - 105.

Gottschling, D. E. (1992). Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. *Proc. Natl. Acad. Sci. USA*. **89**. 4062 - 4065.

Güldener, U., Heck, S., Fiedler, T., Beinhauer, J. and Hegemann, J. H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. *Nuc. Acids. Res.* **24**. 2519 - 2524.

Guthrie, C. and Fink, G. R. [eds.] (1991). Guide to yeast genetics and molecular biology. *Meth. Enzymol.* **194**. Academic Press Inc. San Diego.

Guo, F., Gopaul, D. N. and Van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*. **389**. 40 - 46.

Haber, J. E. (1983). Mating-type genes of *Saccharomyces cerevisiae* in Mobile Genetic Elements. Edited by Shapiro, J. A. Academic Press. London.

Haber, J. E. (1995). *In vivo* biochemistry: physical monitoring of recombination induced by site-specific endonucleases. *BioEssays*. **17**. 609 - 620.

Haber, J. E. (1998a). A locus control region regulates yeast recombination. *Trends In Genetics*. **14**. 317 - 321.

Haber, J. E. (1998b). Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **32**. 561 - 599.

Haber, J. E., Ray, B. L., Kolb, J. M. and White, C. I. (1993). Rapid kinetics of mismatch repair of heteroduplex DNA that is formed during recombination in yeast. *Proc. Natl. Acad. Sci.* **90**. 3363 - 3367.

Haber, J. E., Rogers, D. T. and McCusker, J. H. (1980a). Homothallic conversions of yeast mating-type genes occurs by intrachromosomal recombination. *Cell*. **22**. 277 - 289.

Haber, J. E., Savage, W. T, Raposa, S. M., Weiffenbach, B. and Rowe, L. B. (1980b). Genetic identification of sequences adjacent to the yeast mating-type locus that are essential for transposition of new mating-type alleles. *Proc. Natl. Acad. Sci. USA*. **77**. 2824 - 2828.

Hagen, D. C., Westby, C. A. and Sprague, G. F. (1993). Transcription of alpha-specific genes in *Saccharomyces cerevisiae*: DNA sequence requirement for activity of the coregulator α1. *Mol. Cell. Biol.* **13**. 6866 - 6875.

Harashima, S., Nogi, Y. and Oshima, Y. (1974). The genetic system controlling homothallism genes, *HMα*/*HMα* and *HMa*/*HMa*, in *Saccharomyces* yeasts. *Genetics*. **77**. 639 - 650.

Harayama, S. (1998). Artificial evolution by DNA shuffling. *Trends Biotech.* **16**. 76 - 82.

Hartung, M. and Kisters-Woike, B. (1998). Cre mutants with altered DNA binding properties. *J. Biol. Chem.* **273**. 22884 - 22891.

Hawthorne, D. C. (1963). Directed mutation of the mating-type alleles as an explanation of homothallism in yeast. Abstracts of the Proceedings of the Eleventh International Congress of Genetics. **1**. 34 - 35.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. and Grunstein, M. (1995). Histone H3 and H4 N termini interact with the silent information regulators Sir3 and Sir4 *in vitro*: a model for the formation of heterochromatin in yeast. *Cell*. **80**. 583 - 592.

Hicks, J. B. and Herskowitz, I. (1976). Interconversion of yeast mating-types. Direct observations of the action of the homothallism (*HO*) gene. *Genetics*. **83**. 245 - 258.

Hicks, J. B., Strathern, J. N. and Herskowitz, I. (1977). The cassette model of mating-type interconversion. *In* DNA insertion elements, plasmids, and episomes (Bukhari, A. I., Shapiro, J. A. and Adhya, S. L., eds.). Cold Spring Harbour Laboratory, New York. 457 - 462.

Hicks, J. B., Strathern, J. N. and Klar, A. J. S. (1979). Transposable mating-type genes in *Saccharomyces cerevisiae*. *Nature*. **282**. 478 - 483.

Hill, D. E. (1989). Integrative transformation of yeast using electroporation. *Nuc. Acids. Res.* **17**. 8011.

Hoekstra, M. F., Burbee, D., Singer, J., Mull, E., Chiao, E. and Heffron, F. (1991). A Tn*3* derivative that can be used to make short in-frame insertions within genes. *Proc. Natl. Acad. Sci. USA*. **88**. 5457 - 5461.

Hoess, R. H., Ziese, M. and Sternberg, N. (1982). P1 site-specific recombination: Nucleotide sequence of the recombining sites. *Proc. Natl. Acad. Sci. USA*. **79**. 3398 - 3402.

Hoess, R. H. and Abremski, K. (1984). Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl. Acad. Sci. USA*. **81**. 1026 - 1029.

Hoess, R. H. and Abremski, K. (1985). Mechanism of strand cleavage and exchange in the Cre-*lox* site-specific recombination system. *J. Mol. Biol.* **181**. 351 - 362.

Hoess, R. H., Wierzbicki, A. and Abremski, K. (1986). The role of the *loxP* spacer region in P1 site-specific recombination. *Nucleic Acids Res.* **14**. 2287 - 2300.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genet. Res.* **5**. 282 - 304.

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. and Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene.* **77**. 61 – 68.

Hunter, N., Chambers, S. R., Louis, E. J. and Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO Journal.* **15**. 1726 - 1733.

Ikeda, H. and Tomizawa, J. (1968). Prophage P1, an extrachromosomal element. *Cold Spring Harbour Symp. Quant. Biol.* **33**. 791 - 798.

Ivanov, E. L., Sugawara, N., White, C. I., Fabre, F. and Haber, J. E. (1994). Mutations in *XRS2* and *RAD50* delay but do not prevent mating-type switching in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14**. 3414 - 3425.

Jacquier, A. and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell.* **41**. 383 - 394.

Joo, H., Lin, Z. and Arnold, F. H. (1999). Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature*. **399**. 670 - 673.

Kayne, P. S., Kim, U. J., Han, M., Yoshizaki, F. and Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating-type loci in yeast. *Cell.* **55**. 27 - 39.

Keleher, C. A., Passmore, S. and Johnson, A. D. (1989). Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. *Mol. Cell. Biol.* **9**. 5228 - 5230.

Klar, A. J. S., Fogel, S. and MacLeod, K. (1979). *MAR1* a regulator of *HM***a** and *HM*α loci in *Saccharomyces cerevisiae*. *Genetics*. **93**. 37 - 50.

Klar, A. J. S., Hicks, J. B. and Strathern, J. N. (1982). Directionality of yeast mating-type interconversion. *Cell*. **28**. 551 - 561.

Klar, A. J. S. and Strathern, J. N. (1984). Resolution of recombination intermediates generated during yeast mating-type switching. *Nature*. **310**. 744 - 748.

Kolkman, J. A. and Stemmer, W. P. C. (2001). Directed evolution of proteins by exon shuffling. *Nature Biotechnology*. **19**. 423 - 427.

Kostriken, R. and Heffron, F. (1984). The product of the *HO* gene is a nuclease: purification and characterisation of the enzyme. *Cold Spring Harbour Symp. Quant. Biol.* **49**. 89 - 97.

Kostriken, R., Strathern, J. N., Klar, A. J. S, Hicks, J. B. and Heffron, F. (1983). A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell*. **35**. 167 - 174.

Kozac, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene*. **234**. 187 - 208.

Kramer, B., Kramer, W., Williamson, M. and Fogel, S. (1989). Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional *PMS* genes. *Mol. Cell. Biol.* **9**. 4432 - 4440.

Kramer, W., Kramer, B., Williamson, M. S. and Fogel, S. (1989). Cloning and nucleotide sequence of DNA mismatch repair gene *PMS1* from *Saccharomyces cerevisiae*: Homology of PMS1 to prokaryotic MutL and HexB. *J. Bacteriol.* **171**. 5339 - 5346.

Krautwurst, D., Yau, K. W. and Reed, R. R. (1998). Identification of ligands for olfactory receptors by functional expression of receptor library. *Cell.* **95**. 917 - 926.

Kurtz, S. and Shore, D. (1991). RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes. Dev.* **5**. 5264 - 5268.

Kwon, H. J., Tirumalai, R., Landy, A. and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science.* **276**. 126 - 131.

Kyrion, G., Liu, K. and Lustig, A. J. (1993). RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes. Dev.* **7**. 1146 - 1159.

Längle-Rouault, F. and Jacobs, E. (1995). A method for performing precise alterations in the yeast genome using a recyclable selectable marker. *Nucleic Acids Res.* **23**. 3079 - 3081.

Lasko, M., Sauer, B., Mosinger, Jr., B., Lee, E. J., Manning, R. W., Yu, S. H., Mulder, K. L. and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl. Acad. Sci. USA.* **89**. 6232 - 6236.

Le, Y., Gagnete, S., Tombaccini, D., Bethke, B. and Sauer, B. (1999). Nuclear targeting determinants of the phage P1 Cre DNA recombinase. *Nucleic. Acids. Res.* **24**. 4703 - 4709.

Lewin, B. (1994). Genes V. Oxford University Press.

Lin, Z. L., Thorsen, T. and Arnold, F. H. (1999). Functional expression of horseradish peroxidase in *E. coli* by directed evolution. *Biotechnol. Prog.* **15**. 467 - 471.

Lindegren, C. C. and Lindegren, G. (1943). A new method for hybridizing yeast. *Proc. Natl. Acad. Sci. USA.* **29**. 306 - 308.

Lohr, D., Venkov, P. and Zlatanova, J. (1995). Transcriptional regulation in the yeast *GAL* gene family: a complex genetic network. *FASEB Journal.* **9**. 777 - 787.

Mack, A., Sauer, B., Abremski, K. and Hoess, R. (1992). Stoichiometry of the Cre recombinase bound to the *lox* recombining site. *Nucleic Acids. Res.* **20**. 4451 - 4455.

MacKay, V. L., Welch, S. K., Insley, M. Y., Manney, T. R., Holly, J., Saari, G. C. and Parker, M. L. (1988). The *Saccharomyces cerevisiae* BAR1 gene encodes an exported protein with homology to pepsin. *Proc. Natl. Acad. Sci. USA.* **85**. 55 - 59.

Macreadie, I. G., Scott, R. M., Zinn, A. R. and Butow, R. A. (1985). Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. *Cell.* **41**. 395 - 402.

Manivasakam, P. and Schiestl, R. H. (1993). High efficiency transformation of *Saccharomyces cerevisiae* by electroporation. *Nuc. Acids. Res.* **21**. 4414 - 4415.

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J, Griffiths, A. D. and Winter, G. (1991). By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**. 581 - 597.

Matsuura, T., Miyai, K., Trakulnaleamsai, S., Yomo, T., Shima, Y., Miki, S., Yamamoto, K. and Urabe, I. (1999). Evolutionary molecular engineering by random elongation mutagenesis. *Nature Biotechnology.* **17**. 58 - 61.

McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*. **348**. 552 - 554.

McGill, C., Shafer, B. and Strathern, J. N. (1989). Coconversion of flanking sequences with homothallic switching. *Cell*. **57**. 459 - 467.

McPherson, M. J., Quirke, P. and Taylor, G. R. [eds.] (1992). *PCR - A practical approach.* Oxford University Press.

Miller, A. M. and Nasmyth, K. (1984). Role of DNA replication in the repression of silent mating-type loci in yeast. *Nature*. **312**. 247 - 251.

Mombaerts, P. (1999). Odorant receptor genes in humans. *Current Opinion in Genetics and Development*. **9**. 315 - 320.

Moore, J. C. and Arnold, F. H. (1996). Directed evolution of a *para*-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology*. **14**. 458 - 467.

Moretti, P., Freeman, K., Coodly, L and Shore, D. (1994). Evidence that a complex of *SIR* proteins interacts with the silencer and telomere binding protein RAP1. *Genes Dev.* **8**. 2257 - 2269.

Mortimer, R. H. and Hawthorne, D. C. (1969). Yeast Genetics. In The Yeasts. (Rose, A. H. and Harrison, J. S., eds.). New York Academic Press. 385 - 460.

Mortimer, R. K. and Johnston, J. R. (1986). Genealogy of principal strains of the Yeast Genetic Stock Centre. *Genetics.* **113**. 35 - 43.

Moazed, D., Kistler, A., Axelrod, A., Rine, J. and Johnson, A. D. (1997). Silent information regulator protein complexes in *Saccharomyces cerevisiae*: a *SIR2/SIR4* complex and evidence for a regulatory domain in *SIR4* that inhibits its interaction with *SIR3*. *Proc. Natl. Acad. Sci. USA.* **94**. 2186 - 2191.

Murray, A. and Hunt, T. (1993). The cell cycle, an introduction. Oxford University Press.

Nasmyth, K. A. and Tatchell, K. (1980). The structure of transposable yeast mating-type loci. *Cell.* **19**. 753 - 764.

Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C and Smith, M. (1981). A position effect in the control of transcription at yeast mating-type loci. *Nature.* **289**. 244 - 250.

Nasmyth, K. (1982). The regulation of yeast mating-type chromatin structure by *SIR*: an action at a distance affecting both transcription and transposition. *Cell.* **30**. 567 - 578.

Nasmyth, K. (1983). Molecular analysis of cell lineage. *Nature.* **302**. 670 - 676.

Nasmyth, K. and Shore, D. (1987). Transcriptional regulation of the yeast life cycle. *Science.* **237**. 1162 - 1170.

Naumov, G. I. and Tolstorukov, I. I. (1973). Comparative genetics of yeast. X. Reidentification of mutators of mating-types in *Saccharomyces*. *Genetika.* **9**. 82 - 91.

Nickoloff, J. A., Chen, E. Y. and Heffron, F. (1986). A 24-base-pair DNA sequence from the *MAT* locus stimulates intergenic recombination in yeast. *Proc. Natl. Acad. Sci. USA.* **83**. 7831 - 7835.

Nickoloff, J. A., Singer, J. D., Hoekstra, M. F. and Heffron, F. (1989). Double-strand breaks stimulate alternative mechanisms of recombination repair. *J. Mol. Biol.* **207**. 527 - 541.

Niedenthal, R. K., Riles, L., Johnston, M. and Hegemann, J. H. (1996). Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast.* **12**. 773 - 786.

O'Gorman, S., Fox, D. T. and Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*. **251**. 1351 - 1355.

Oliver, S. G. *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature*. **357**. 38 - 46.

Orr-Weaver, T. L., Szostak, J. W. and Rothstein, J. R. (1981). Yeast transformation: A model system for the study of recombination. *Proc. Natl. Acad. Sci. USA*. **78**. 6354 - 6358.

Ostermeier, M., Nixon, A. E. and Benkovic, S. J. (1999a). Incremental truncation as a strategy in the engineering of novel biocatalysts. *Bioorg. Med. Chem.* **7**. 2139 - 2144.

Ostermeier, M., Shim, J. H. and Benkovic, S. J. (1999b). A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology*. **17**. 1205 - 1210.

Paques, F. and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**. 349 - 404.

Pays, E. and Nolan, D. P. (1998). Expression and function of surface proteins in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **91**. 3 - 36.

Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P. and Fodor, S. P. A. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA*. **91**. 5022 - 5026.

Plessis, A., Perrin, A., Haber, J. E. and Dujon, B. (1992). Site-specific recombination determined by I-*Sce*I mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics*. **130**. 451 - 460.

Puchta, H., Dujon, B. and Hohn, B. (1993). Homologous recombination in plant cells is enhanced by *in vivo* induction of double strand breaks into DNA by a site-specific endonuclease. *Nuc. Acids. Res.* **21**. 5034 - 5040.

Qin, M., Bayley, C., Stockton, T. and Ow, D. W. (1994). Cre recombinase-mediated site-specific recombination between plant chromosomes. *Proc. Natl. Acad. Sci. USA.* **91**. 1706 - 1710.

Ramsay, G. (1998). DNA chips: State-of-the art. *Nature Biotechnology.* **16**. 40 - 48.

Raveh, D., Hughes, S. H., Shafer, B. K. and Strathern, J. N. (1989). Analysis of the HO-cleaved *MAT* DNA intermediate generated during the mating-type switch in the yeast *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **220**. 33 - 42.

Ravindra, A., Weiss, K. and Simpson, R. T. (1999). High-resolution structural analysis of chromatin at specific loci: *Saccharomyces cerevisiae* silent mating-type locus *HMR***a**. **19**. 7944 - 7950.

Ray, A., Siddiqi, I., Kolodkin, A. L. and Stahl, F. W. (1988). Intra-chromosomal gene conversion induced by a DNA double-strand break in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **201**. 247 - 260.

Ray, B. L., White, C. I. and Haber, J. E. (1991). Heteroduplex formation and mismatch repair of the 'stuck' mutation during mating-type switching in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**. 5372 - 5380.

Rayssiguier, C., Thaler, D. S. and Radman, M. (1989). The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*. **342**. 396 - 401.

Redfern, M. R. (2000). Selection of biologically active peptides using the yeast, *Saccharomyces cerevisiae*. Ph. D. Thesis. University of Bath.

Reenan, R. A. and Kolodner, R. D. (1992). Isollation and characterisation of two *Saccharomyces cerevisiae* genes encoding homologues of the bacterial HexA and MutS mismatch repair proteins. *Genetics*. **132**. 963 - 973.

Reidhaar-Olson, J., Bowie, J. U., Breyer, R. M., Hu, J. C., Knight, K. L., Lim, W. A., Mossing, M. C., Parsell, D. A, Shoemaker, K. R. and Sauer, R. T. (1991). Random mutagenesis of protein sequences using oligonucleotide cassettes. *Methods Enzymol.* **208**. 564 - 603.

Resnick, M. A. and Martin, P. (1976). The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control. *Mol. Gen. Genet.* **143**. 119 - 129.

Resnick, M. A., Zgaga, Z., Hieter, P., Westmoreland, J., Fogel, S and Nilsson-Tillgren, T. (1992). Recombinational repair of diverged DNAs: a study of homeologous chromosomes and mammalian YACs in yeast. *Mol. Gen. Genet.* **234**. 65 - 73.

Restrepo, B. I. and Barbour, A. G. (1994). Antigen diversity in the bacterium *B. hermsii* through 'somatic' mutations in rearranged *vmp* genes. *Cell*. **78**. 867 - 876.

Reynaud, C. A., Anquez, V., Grimal, H. and Weill, J. C. (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell*. **48**. 379 - 388.

Ricchetti, M., Fairhead, C. and Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature*. **402**. 96 - 100.

Rine, J. D., Strathern, J. B., Hicks, J. B. and Herskowitz, I. (1979). A suppressor of mating-type locus mutations in *Saccharomyces cerevisiae*: evidence for and identification of cryptic mating-type loci. *Genetics*. **93**. 877 - 901.

Roman, H. and Sands, S. M. (1953). Heterogeneity of clones of *Saccharomyces* derived from haploid ascospores. *Proc. Natl. Acad. Sci. USA*. **39**. 171 - 179.

Romanos, M. A., Scorer, C. A., Clare, J. J. (1992). Foreign gene expression in yeast: a review. *Yeast*. **8**. 423 - 488.

Rong, Y. S. and Golic, K. G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science*. **288**. 2013 - 2018.

Ross-MacDonald, P., Sheehan, A., Roeder, G. S. and Snyder, M. (1997). A multi-purpose transposon system for analysing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*. **94**. 190 - 195.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: Integrative DNA transformation in yeast. *Methods Enzymol.* **194**. 281 - 301.

Rouet, P., Smih, F. and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**. 8096 - 8106.

Sambrook, J., Fritsch, E. F. and Maniatis, T. [eds.] (1989). *Molecular cloning. A laboratory manual.* Cold Spring Harbour Laboratory Press.

Sanger, F., Nicken, S. and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**. 5463 - 5467.

Santos-Rosa, H. and Aguilera, A. (1994). Increase in incidence of chromosome instability and non-conservative recombination between repeats in *Saccharomyces cerevisiae hpr1*Δ strains. *Mol. Gen. Genet.* **245**. 224 - 236.

Sauer, B. (1987). Functional expression of the *cre-lox* site-specific recombination system in the yeast *Saccharomyces cerevisiae. Mol. Cell. Biol.* **7**. 2087 - 2096.

Sauer, B. and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci. USA.* **85**. 5166 - 5170.

Sauer, B. and Henderson, N. (1989). Cre-stiimulated recombination at *loxP*-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* **17**. 147 - 161.

Sauer, B. (1992). Identification of cryptic *lox* sites in the yeast genome by selection for *Cre*-mediated chromosome translocations that confer multiple drug resistance. *J. Mol. Biol.* **223**. 911 - 927.

Sauer, B. (1993). Manipulation of transgenes by site-specific recombination: Use of Cre recombinase. *Meth. Enzymol.* **225**. 890 - 900.

Sauer, B. (1996). Multiplex Cre/*lox* recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. *Nucleic Acids Res.* **24**. 4608 - 4613.

Scherer, S. and Davis, R. W. (1979). Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. USA.* **76**. 4951 - 4960.

Schnell, R. and Rine, J. (1986). A position effect on the expression of a tRNA gene mediated by the *SIR* genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**. 494 - 501.

Selva, E. M., New, L., Crouse, G. F. and Lahue, R. S. (1995). Mismatch correction acts as a barrier to homeologous recombination in *Saccharomyces cerevisiae*. *Genetics*. **139**. 1175 - 1188.

Senecoff, J. F., Bruckner, R. C. and Cox, M. M. (1985). The FLP recombinase of the yeast 2-μm plasmid: characterisation of its recombination site. *Proc. Natl. Acad. Sci. USA*. **82**. 7270 - 7274.

Shao, Z., Zhao, H., Giver, L. and Arnold, F. H. (1998). Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nuc. Acids Res.* **26**. 681 - 683.

Shei, G. J. and Broach, J. R. (1995). Yeast silencers can act as orientation-dependent gene inactivation centres that respond to environmental signals. *Mol. Cell. Biol.* **15**. 3496 - 3506.

Shen, P. and Huang, H. V. (1989). Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* **218.** 358 - 360.

Sieber, V., Martinez, C. A. and Arnold, F. H. (2001). Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnology*. **19**. 456 - 460.

Sikorski, R. and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. **122**. 19 - 27.

Smith, G. P. (1994). The progeny of sexual PCR. *Nature*. **370**. 324 - 325.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by agarose gel electrophoresis. *J. Mol. Biol.* **98**. 503 - 507.

Stemmer, W. P. C. (1994a). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*. **370**. 389 - 391.

Stemmer, W. P. C. (1994b). DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA*. **91**. 10747 - 10751.

Sternberg, N. and Austin, S. (1981). The maintenance of the P1 plasmid prophage. *Plasmid*. **5**. 20 - 31.

Sternberg, N. and Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. I. Recombination between *loxP* sites. *J. Mol. Biol.* **150**. 467 - 486.

Sternberg, N., Hamilton, D. and Hoess, R. (1981). Bacteriophage P1 site-specific recombination. II. Recombination between *loxP* and the bacterial chromosome. *J. Mol. Biol.* **150**. 487 - 507.

Sternberg, N., Sauer, B., Hoess, R. and Abremski, K. (1986). Bacteriophage P1 *cre* gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. *J. Mol. Biol.* **187**. 197 - 212.

Stoenner, H. G., Dodd, T., and Larsen, C. (1982). Antigenic variation of *Borrelia hermsii*. *J. Exp. Med.* **156**. 1279 - 1311.

Strahl-Bolsinger, S., Hecht, A., Luo, K. and Grunstein, M. (1997). *SIR2* and *SIR4* interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**. 83 - 93.

Strathern, J. N. and Herskowitz, I. (1979). Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. *Cell.* **17**. 371 - 381.

Strathern, J. N., Hicks, J. B. and Herskowitz, I. (1981). Control of cell type in yeast by the mating-type locus: the **α**1-**α**2 hypothesis. *J. Mol. Biol.* **147**. 357 - 373.

Strathern, J. N., Klar, A. J. S., Hicks, J. B., Abraham, J. A., Nasmyth, K. A. and McGill, C. (1982). Homothallic switching of yeast mating-type cassettes is initiated by a double-stranded cut in the *MAT* locus. *Cell.* **31**. 183 - 192.

Strathern, J. N., Spatola, E., McGill, C. and Hicks, J. B. (1980). Structure and organisation of transposable mating-type cassettes in *Saccharomyces* yeasts. *Proc. Natl. Acad. Sci. USA.* **77**. 2839 - 2843.

Sugawara, N. and Haber, J. E. (1992). Characterisation of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Mol. Cell. Biol.* **12**. 563 - 575.

Sugawara, N., Ivanov, E. L., Fishman-Lobel, J., Ray, B. L., Wu, X. and Haber, J. E. (1995). DNA structure-dependent requirements for yeast *RAD* genes in gene conversion. *Nature*. **373**. 84 - 86.

Szeto, L., Fafalios, M. K., Zhong, H., Vershon, A. K. and Broach, J. R. (1997). α2p controls donor preference during mating-type interconversion in yeast by inactivating a recombinational enhancer on chromosome III. *Genes Dev.* **11**. 1899 - 1911.

Szostak, J, W., Orr-Weaver, T. L. and Rothstein, R. J. (1983). The double-strand-break-repair model for recombination. *Cell*. **33**. 25 - 35.

Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. and Schmitz, G. (1990). Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-*Sce*I. *Nuc. Acids. Res.* **19**. 189 - 190.

Thierry, A. and Dujon, B. (1992). Nested chromosomal fragmentation in yeast using the meganuclease I-*Sce*I: a new method for physical mapping of eukaryotic genomes. *Nuc. Acids. Res.* **20**. 5625 - 5631.

Thomson, J. S., Hecht, A. and Grunstein, M. (1994). The histone H3 amino terminus is required for both telomeric and silent mating locus repression in yeast. *Nature*. **369**. 235 - 247.

Tomlinson, I. M., Walter, G. and Marks, J. D. (1992). The repertoire of human germline V(H) sequences reveals about 50 groups of V(H) segments with different hypervariable loops. *J. Mol. Biol.* **227**. 776 – 798.

Triolo, T. and Sternglanz, R. (1996). Role of interactions between the origin recognition complex and *SIR1* in transcriptional silencing. *Nature*. **381**. 251 - 253.

Tsubouchi, H. and Ogawa, H. (1998). A novel *mre11* mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. *Mol. Cell. Biol.* **18**. 260 - 68.

Tsurushita, N., Fu, H. and Warren, C. (1996). Phage display vectors for *in vivo* recombination of immunoglobulin heavy and light chain genes to make large combinatorial libraries. *Gene*. **172**. 59 - 63.

van der Ploeg, L. H. T., Gottesdiener, K. and Lee, M. G. S. (1992). Antigenic variation in African trypanosomes. *Trends Genet*. **8**. 452 - 457.

van Deursen, J., Fornerod, M., van Rees, B. and Grosveld, G. (1995). Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc. Natl. Acad. Sci. USA*. **92**. 7376 - 7380.

Wach, A., Brachat, A., Pohlmann, R. and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruption in *Saccharomyces cerevisiae*. *Yeast*. **10**. 1793 – 1808.

Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast*. **12**. 259 - 256.

Wach, A., Brachat, A. and Philippsen, P. (1996). Guidelines for EUROFAN B0 programme: ORF deletants, plasmid tools, basic functional analyses. Institute for Applied Microbiology, University of Basel, Switerland.

Waterhouse, P., Griffiths, A. D., Johnson, K. S. and Winter, G. (1993). Combinatorial infection and *in vivo* recombination: a strategy for making large phage antibody repertoires. *Nucleic Acids Res.* **21**. 2265 - 2266.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. (1987). Molecular biology of the gene. Benjamin/Cummings Publishing Company. 580.

Weaver, R. F. and Hedrick, P. W. (1992). Genetics. Second edition. Wm. C. Brown Publishers.

Weiss, K. and Simpson, R. T. (1997). Cell type-specific chromatin organisation of the region that governs directionality of yeast mating-type switching. *EMBO Journal.* **16**. 4352 - 4360.

White, C. I. and Haber, J. E. (1990). Intermediates of recombination during mating-type switching in *Saccharomyces cerevisiae*. *EMBO Journal.* **9**. 663 - 673.

Williamson, M. S., Game, J. C. and Fogel, F. (1985). Meiotic gene conversion mutants in *Saccharomyces cerevisiae*. Isolation and characterisation of *pms1-1* and *pms1-2*. *Genetics.* **110**. 609 - 646.

Winge, O. and Hjort, A. (1935). On some *Saccharomycetes* and other fungi still alive in the pure cultures of Emil Chr. Hansen and Alb. Klocker. C. R. *Trav. Lab. Carlsnerg Ser. Physiol.* **21**. 51 - 58.

Winston, F., Dollard, C. and Ricupero-Hovasse, S. L. (1995). Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast.* **11**. 53 – 55.

Winter, G. and Milstein, C. (1991). Man-made antibodies. *Nature*. **349**. 293 - 299.

Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**. 433 - 455.

Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. and Lockhart, D. J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnology*. **15**. 1359 - 1367.

Wu, X. and Haber, J. E. (1995). *MAT***a** donor preference in yeast mating-type switching: activation of a large chromosomal region for recombination. *Genes Dev.* **9**. 1922 - 1932.

Wu, X. and Haber, J. E. (1996). A 700 b.p. *cis*-acting region controls mating-type dependent recombination along the entire left arm of yeast chromosome III. *Cell.* **87**. 277 - 285.

Wu, C., Weiss, K., Yang, C., Harris, M. A., Bik-Kwoon, T., Newlon, C. S., Simpson, R. T. and Haber, J. E. (1998). Mcm1 regulates donor preference controlled by the recombination enhancer in *Saccharomyces* mating-type switching. *Genes Dev.* **12**. 1726 - 1737.

Wu, X., Wu, C. and Haber, J. E. (1997). Rules of donor preference in *Saccharomyces* mating-type gene switching revealed by a competition assay involving two types of recombination. *Genetics.* **147**. 399 - 407.

You, L. and Arnold, F. (1994). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**. 77 - 83.

Zhang, J., Dawes, G. and Stemmer, W. P. C. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. USA.* **94**. 4504 - 4509.

Zhao, H. and Arnold, F. H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**. 47 - 53.

Zhao, H., Giver, L., Shao, Z., Affholter, J. A. and Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology.* **16**. 258 - 261.

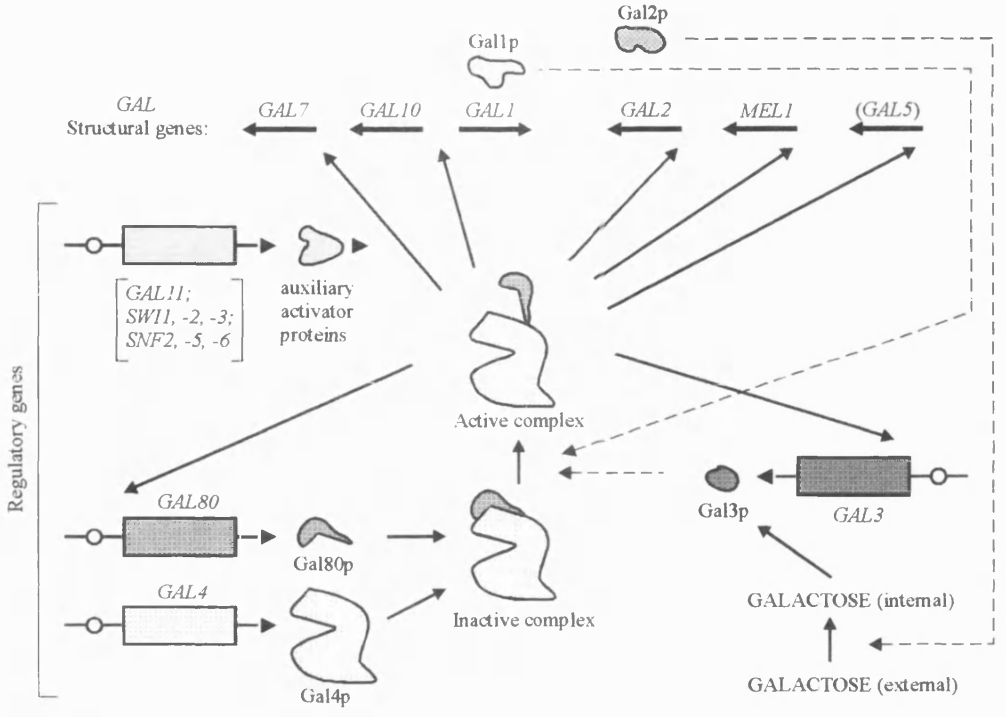
APPENDIX

A. The *GAL1* inducible promoter

The nature and origin of the *GAL1* inducible promoter is described briefly here, as it is used to drive expression of the I-*Sce*I endonuclease (Chapter 3), and for expression of the *Cre*-recombinase (Chapter 4).

The *GAL* gene family is one of the best-studied systems in *S. cerevisiae*, providing a major model for the study of eukaryotic regulation (Lohr *et al.*, 1995). The family is comprised of structural and regulatory genes that enable cells to utilise galactose as a carbon source. The structural genes are *GAL2, GAL1, GAL7, GAL10, MEL1* and *GAL5*, their gene products transport galactose into cells, convert intracellular galactose to the glycolytic substrate glucose-1-phosphate, and provide a galactosidase activity. The genes are strongly regulated at the transcriptional level by the carbon source, three main types of regulated states are observed: inactive-repressed (glucose); inactive, poised for induction (raffinose); active, induced to high level expression (galactose).

The specific regulatory genes *GAL3*, *GAL4*, and *GAL80*, provide the major structural gene regulators. Gal4p is required for the galactose-induction of *GAL* gene expression. The protein binds via its amino-terminus to sequences located upstream (~100 to 400 b.p.) from each of the structural genes [upstream activation sequences (UAS)] and activates transcription through a specific carboxy-terminal domain. Gal80p binds to and masks the Gal4p transcription activation function in non-galactose carbon sources. Gal3p mediates the galactose-induced release of Gal80p inhibition of Gal4p. The regulatory strategy is shown diagrammatically below.

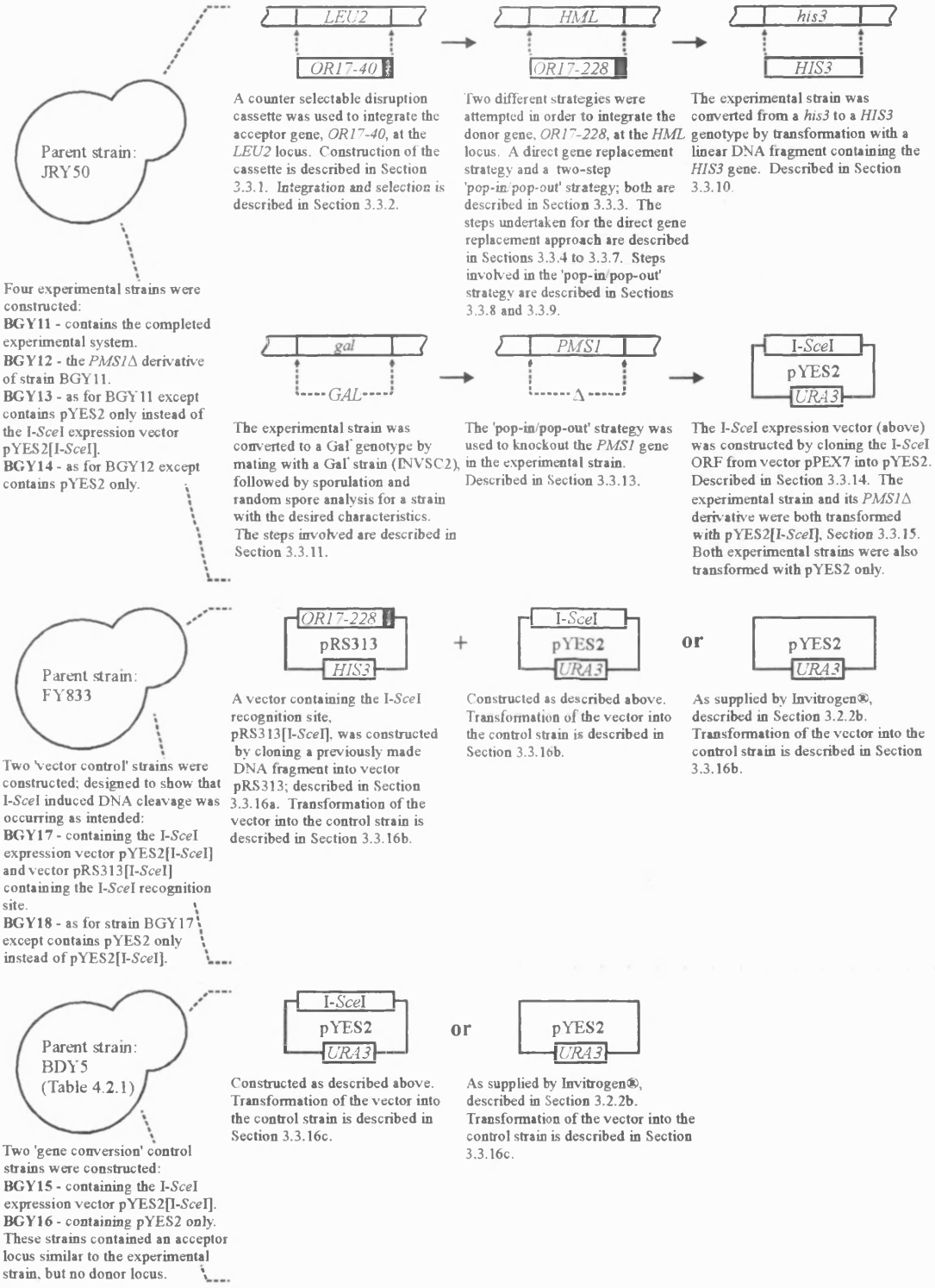


Galactose induction pathways. The regulatory strategy is implemented by a network of activating and repressing activities. *GAL* structural gene are at the top, large arrows denote the transcription units for the various genes. Regulatory genes are shown on the left and right sides. Arrows (thin, solid or dotted) refer to the regulation flow within the family. Reproduced from Lohr *et al.* (1995).

The mechanisms that activates the genes is extremely powerful. Induction in galactose results in a ×1000 fold increase in expression levels (*GAL1*, *-7* and *-10*). In glucose or raffinose *GAL1*, *-2*, *-7* and *-10* are not detectably expressed, so the mechanisms that inhibit their expression under these conditions must also be highly effective.

The UAS (the Gal4p binding site) can confer the same levels of galactose inducibility on a heterologous gene, a property that has been widely used for heterologous expression. In this work the *GAL1* portion of the divergent *GAL1-10* promoter was used to regulate expression via carbon source in the vectors used.

B. Summary of steps taken in the construction of strains for Chapter 3



Summary of the main steps involved in the construction of the experimental and control strains for Chapter 3. The steps shown are in the same order as described in Section 3.3 - Methods: Generation of experimental strains - where each individual step is described in detail. The final completed experimental and control strains are briefly described on the left hand side. Detailed descriptions of the genotype of these strains, and of intermediate strains made, are given in Table 3.2.1. The I-*Sce*I recognition site is represented by a filled rectangle.